Tutorial T15

AMIA Fall Symposium
Saturday, November 3, 2001

Customizing the UMLS Metathesaurus for Your Applications

Olivier Bodenreider, MD, PhD
William T. Hole, MD
Betsy L. Humphreys, MLS
Laura Roth, MLS
Suresh Srinivasan, MS

Outline of Tutorial

- Why customize? — Betsy Humphreys
- Metathesaurus basics — Olivier Bodenreider
- How to customize?
  - Customize sources (MetamorphoSys) — L. Roth & S. Srinivasan
  - Customize strings — Olivier Bodenreider
  - Customize synonyms
  - Customize relationships
  - Customize concept spaces
- Adding “local” terminology — Bill Hole

UMLS Knowledge Sources

Multi-purpose tools or “intellectual middleware” for System Developers

- Metathesaurus
- SPECIALIST lexicon and lexical programs
- Semantic Network

UMLS Metathesaurus

- Concepts, terms, and attributes from many controlled “vocabularies”
- New inter-source relationships, definitional information, use information
- Scope determined by combined scope of source vocabularies

UMLS Source "Vocabularies"

- Widely varying purposes, structures, properties, but all are in essence "sets of valid values" for data elements:
  - Thesauri, e.g., MeSH
  - Statistical Classifications, e.g., ICD
  - Billing Codes, e.g., CPT
  - Clinical coding systems, e.g., SNOMED, Read
  - Lists of controlled terms, e.g., COSTAR, HL7 values
- All HIPAA code sets, except NDC

2001 UMLS Metathesaurus

- ~800,000 concepts
- ~1,500,000 “terms” (Eye, Eyes, eye = 1)
- ~1,700,000 “strings”/concept names - (Eye, Eyes, eye = 3)
- ~10,600,000 relationships between concepts
- >50 source vocabularies (including several “families” with multiple members)

How to combine them?

Meta Processor,
Alpha 0.001

Not really ….

- “The Metathesaurus preserves the meanings, hierarchical connections, and other relationships between terms present in its source vocabularies, while adding certain basic information about each of its concepts and establishing new relationships between concepts and terms from different source vocabularies.”

Why Customize? 3 basic reasons

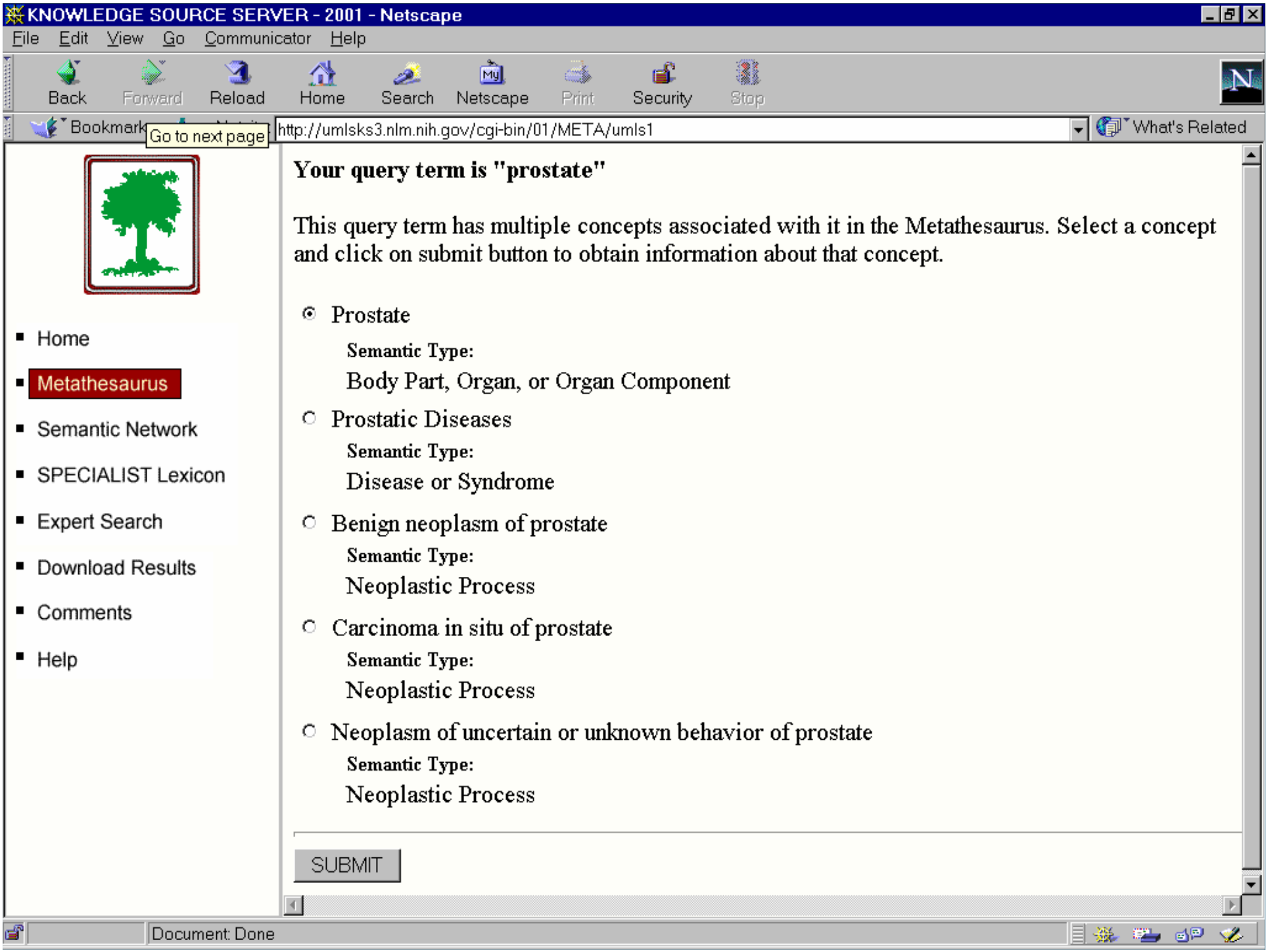
- Because nobody needs or wants all of it for any specific set of purposes
  - extraneous vs. pernicious concepts, strings, relationships
- Because you don't have the licenses required for operational use of all source vocabularies
- Because the default "preferred name" is not best for your applications

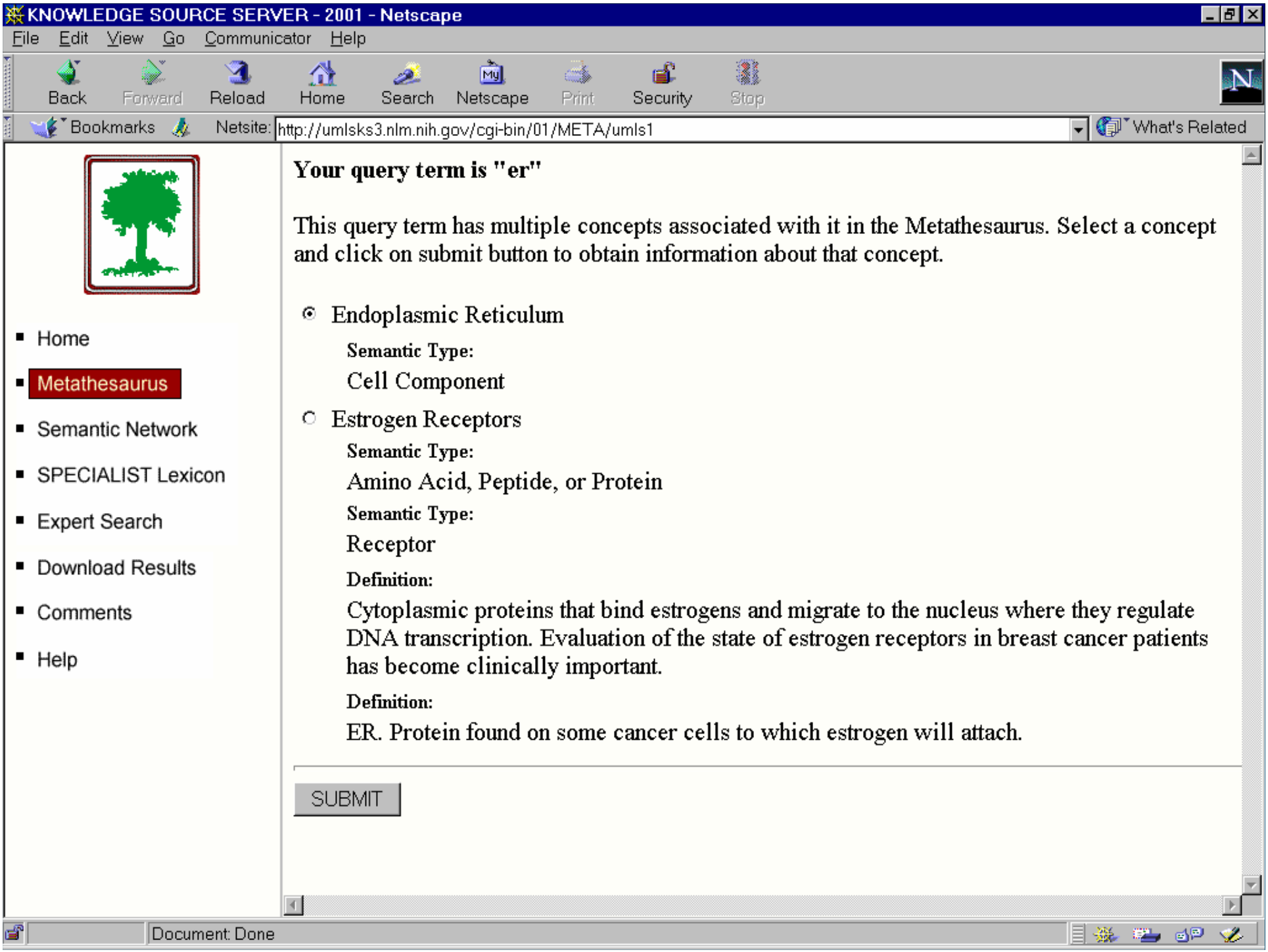
Possibly Extraneous, e.g.,

- Terms in languages other than English
- Redundant minor variations
- Procedure codes, when your application is focused on problems

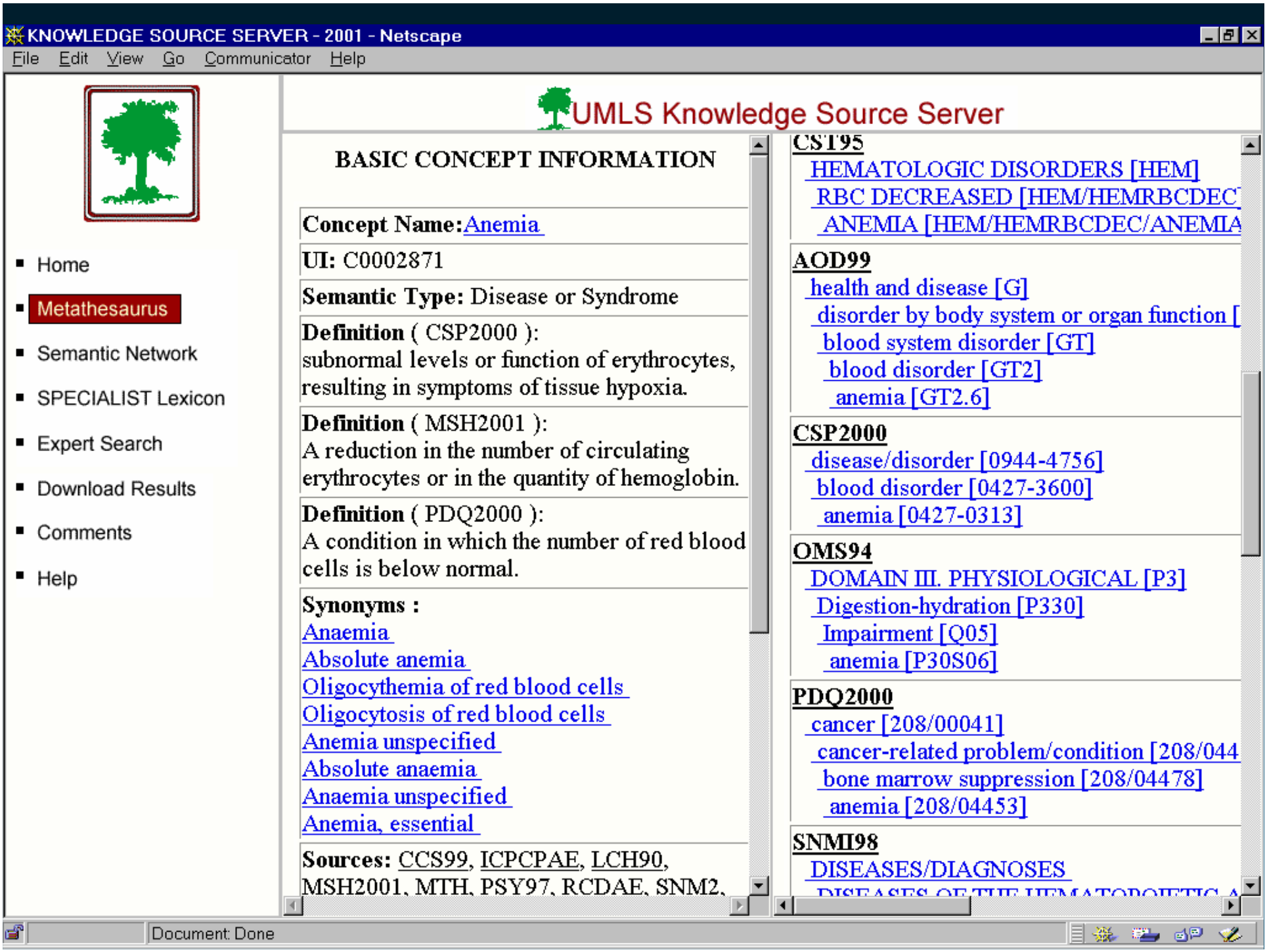
Possibly Pernicious, e.g.,

- Terms that lack face validity
- Abbreviations and short forms
- Other less than beautiful “suppressible synonyms” already identified by NLM
- Relationships that reflect an alien or unhelpful “world view”


KNOWLEDGE SOURCE SERVER - 2001 - Netscape
File Edit View Go Communicator Help
Back Forward Reload Home Search Netscape Print Security Stop
Bookmark
Go to next page
http://umlsks3.nlm.nih.gov/cgi-bin/01/META/umls1
What's Related
Home
Metathesaurus
Semantic Network
SPECIALIST Lexicon
Expert Search
Download Results
Comments
Help
Your query term is "prostate"
This query term has multiple concepts associated with it in the Metathesaurus. Select a concept and click on submit button to obtain information about that concept.
Prostate
Semantic Type:
Body Part, Organ, or Organ Component
Prostatic Diseases
Semantic Type:
Disease or Syndrome
Benign neoplasm of prostate
Semantic Type:
Neoplastic Process
Carcinoma in situ of prostate
Semantic Type:
Neoplastic Process
Neoplasm of uncertain or unknown behavior of prostate
Semantic Type:
Neoplastic Process
SUBMIT
Document: Done


KNOWLEDGE SOURCE SERVER - 2001 - Netscape
File Edit View Go Communicator Help
Back Forward Reload Home Search Netscape Print Security Stop
Bookmarks Netsite: http://umlsks3.nlm.nih.gov/cgi-bin/01/META/umls1
What's Related
Home
Metathesaurus
Semantic Network
SPECIALIST Lexicon
Expert Search
Download Results
Comments
Help
Your query term is "er"
This query term has multiple concepts associated with it in the Metathesaurus. Select a concept and click on submit button to obtain information about that concept.
Endoplasmic Reticulum
Semantic Type:
Cell Component
Estrogen Receptors
Semantic Type:
Amino Acid, Peptide, or Protein
Semantic Type:
Receptor
Definition:
Cytoplasmic proteins that bind estrogens and migrate to the nucleus where they regulate DNA transcription. Evaluation of the state of estrogen receptors in breast cancer patients has become clinically important.
Definition:
ER. Protein found on some cancer cells to which estrogen will attach.
SUBMIT
Document: Done


KNOWLEDGE SOURCE SERVER - 2000 - Netscape
File Edit View Go Communicator Help
Bookmarks Netsite: https://umlsks.nlm.nih.gov/cgi-bin/00/META/umls1
What's Related
Intranet Home P
Welcome to Real
UMLS KNOWLEDGE
KNOWLEDGE SOURC
Meta Help
7 New Publishin
U.S. Government
UMLS Knowledge Source Server
Home
Metathesaurus
Semantic Network
SPECIALIST Lexicon
Expert Search
Download Results
Comments
Help
ALMECILLIN:SUSCEPTIBILITY:POINT IN TIME:ISOLATE AND SERUM:ORDINAL:SERUM BACTERICIDAL TITER
AMDINOCILLIN:SUSCEPTIBILITY:POINT IN TIME:ISOLATE AND SERUM:ORDINAL:SERUM BACTERICIDAL TITER
AMIKACIN:SUSCEPTIBILITY:POINT IN TIME:ISOLATE AND SERUM:ORDINAL:SERUM BACTERICIDAL TITER
AMOXICILLIN:SUSCEPTIBILITY:POINT IN TIME:ISOLATE AND SERUM:ORDINAL:SERUM BACTERICIDAL TITER
AMOXICILLIN AND CLAVULANATE:SUSCEPTIBILITY:POINT IN TIME:ISOLATE AND SERUM:ORDINAL:SERUM BACTERICIDAL TITER
AMPHOTERICIN B:SUSCEPTIBILITY:POINT IN TIME:ISOLATE AND SERUM:ORDINAL:SERUM BACTERICIDAL TITER
AMPICILLIN:SUSCEPTIBILITY:POINT IN TIME:ISOLATE AND SERUM:ORDINAL:SERUM BACTERICIDAL TITER
AMPICILLIN AND SULBACTAM:SUSCEPTIBILITY:POINT IN TIME:ISOLATE AND SERUM:ORDINAL:SERUM BACTERICIDAL TITER
AZITHROMYCIN:SUSCEPTIBILITY:POINT IN TIME:ISOLATE AND SERUM:ORDINAL:SERUM BACTERICIDAL TITER
AZLOCILLIN:SUSCEPTIBILITY:POINT IN TIME:ISOLATE AND SERUM:ORDINAL:SERUM BACTERICIDAL TITER
AZTREONAM:SUSCEPTIBILITY:POINT IN TIME:ISOLATE AND SERUM:ORDINAL:SERUM BACTERICIDAL TITER
BACAMPICILLIN:SUSCEPTIBILITY:POINT IN TIME:ISOLATE AND SERUM:ORDINAL:SERUM BACTERICIDAL TITER
BUTIROSIN:SUSCEPTIBILITY:POINT IN TIME:ISOLATE AND SERUM:ORDINAL:SERUM BACTERICIDAL TITER
Document: Done
Start
GroupW...
Mail Fro...
Mail Fro...
Day
NIH Inte...
KNOW...
Mail Fro...
Microso...
3:10 PM


KNOWLEDGE SOURCE SERVER - 2001 - Netscape
File Edit View Go Communicator Help
UMLS Knowledge Source Server
Home
Metathesaurus
Semantic Network
SPECIALIST Lexicon
Expert Search
Download Results
Comments
Help
BASIC CONCEPT INFORMATION
Concept Name:Anemia
UI: C0002871
Semantic Type: Disease or Syndrome
Definition (CSP2000):
subnormal levels or function of erythrocytes, resulting in symptoms of tissue hypoxia.
Definition (MSH2001):
A reduction in the number of circulating erythrocytes or in the quantity of hemoglobin.
Definition (PDQ2000):
A condition in which the number of red blood cells is below normal.
Synonyms :
Anaemia
Absolute anemia
Oligocythemia of red blood cells
Oligocytosis of red blood cells
Anemia unspecified
Absolute anaemia
Anaemia unspecified
Anemia, essential
Sources: CCS99, ICPCPAE, LCH90, MSH2001, MTH, PSY97, RCDAE, SNM2,
CST95
HEMATOLOGIC DISORDERS [HEM]
RBC DECREASED [HEM/HEMRBCDEC]
ANEMIA [HEM/HEMRBCDEC/ANEMIA]
AOD99
health and disease [G]
disorder by body system or organ function [
blood system disorder [GT]
blood disorder [GT2]
anemia [GT2.6]
CSP2000
disease/disorder [0944-4756]
blood disorder [0427-3600]
anemia [0427-0313]
OMS94
DOMAIN III. PHYSIOLOGICAL [P3]
Digestion-hydration [P330]
Impairment [Q05]
anemia [P30S06]
PDQ2000
cancer [208/00041]
cancer-related problem/condition [208/044
bone marrow suppression [208/04478]
anemia [208/04453]
SNMI98
DISEASES/DIAGNOSES
Document: Done

License restriction levels

- Level 0 - 56.1% of concepts
  - Basic license requirements, e.g., copyright statement and credits to NLM and producers of the vocabularies you use, no redistribution except as a part of your application
- Level 1 - 5.5% of concepts
  - Basic, plus you must negotiate with producer to translate into another language

READ the license, including the appendix

License restriction levels

- Level 2 - 0.1% of concepts
  - Basic, plus you must negotiate with producer for use in the creation of health data
- Level 3 - 38.2% of concepts
  - Basic, plus you must negotiate with the producer for *any* production use. Explicit prohibition against providing access via the Internet.
- There may - or may not - be license fees associated with uses not covered by the UMLS license.

READ the license, including the appendix

NATIONAL LIBRARY OF MEDICINE

Customization is critical,

but it ***requires*** a clear understanding of:

- Your functional requirements
- Characteristics of relevant UMLS source vocabularies
  - You can explore these via UMLS Knowledge Source Server
- Your license arrangements
- -- ***and*** Technical expertise
- Therefore, it is usually a team sport.

Outline of Tutorial

- Why customize? — Betsy Humphreys
- Metathesaurus basics — Olivier Bodenreider
- How to customize?
  - Customize sources (MetamorphoSys) — L. Roth & S. Srinivasan
  - Customize strings — Olivier Bodenreider
  - Customize synonyms
  - Customize relationships
  - Customize concept spaces
- Adding “local” terminology — Bill Hole

Access to UMLS data

- Local database
- Data model
  - Relational model + SQL
  - Object-oriented model + some O-O language

Metathesaurus Basic organization

- Synonymous terms clustered into a concept
- Preferred term (default)
- Unique identifier (CUI)

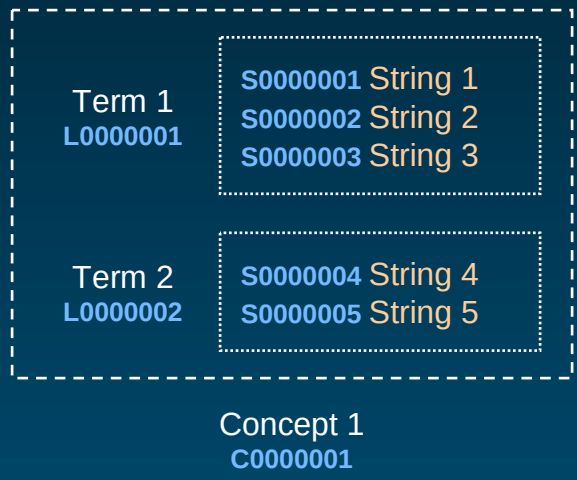
Adrenal gland diseases	MeSH	D000307
Adrenal disorder	AOD	0000005418
Disorder of adrenal gland	Read	C15z.
Diseases of the adrenal glands	SNOMED	DB-70000

C0001621

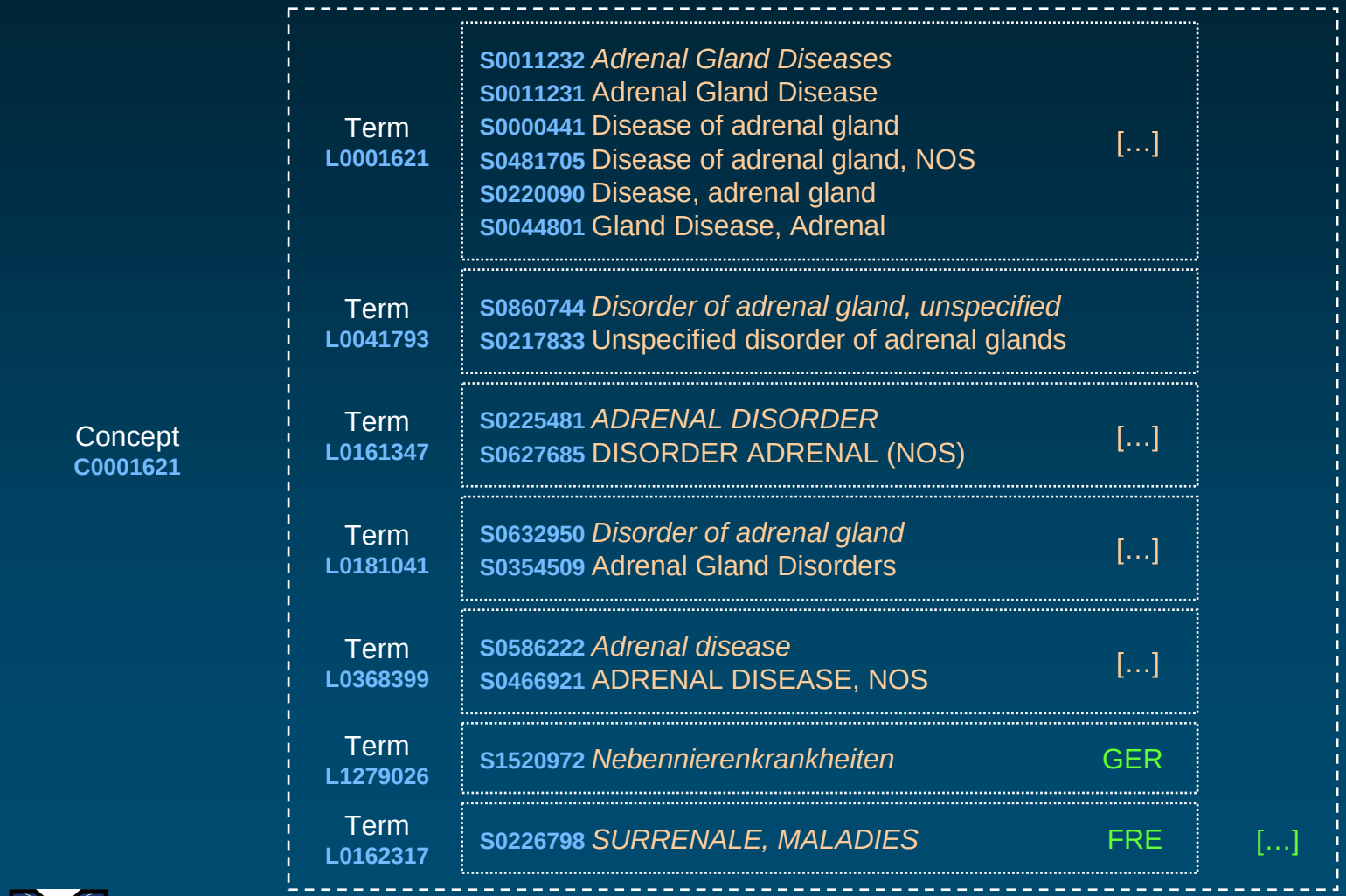
Adrenal Gland Diseases

Metathesaurus Concepts

- Concept: Cluster of synonymous terms
  - ~800,000 concepts
  - identified by a CUI
- Term: Set of lexical variants
  - ~1.5 M terms
  - identified by a LUI
- String: Concept name
  - ~1.7 M strings
  - identified by a SUI



Cluster of synonymous terms



Metathesaurus files Concepts

MRCON

Adrenal gland diseases
Adrenal disorder
Disorder of adrenal gland
Diseases of the adrenal glands

C0001621

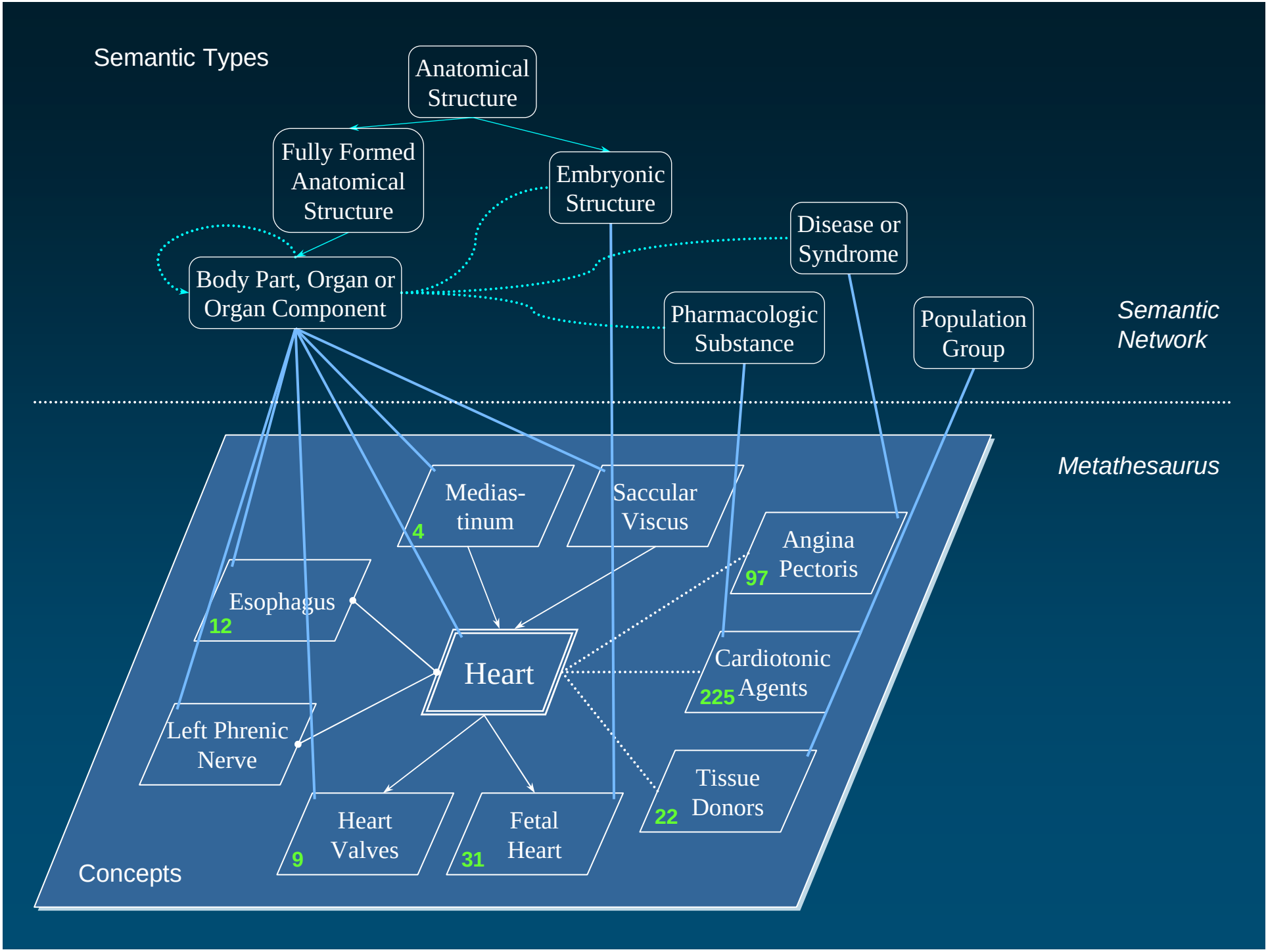
MRSO

MeSH	D000307
AOD	0000005418
Read	C15z.
SNOMED	DB-70000

Adrenal Gland Diseases

Metathesaurus Relationships

- Asserted relationships: 4.7 M pairs of concepts
- Statistical relationships : 5.9 M pairs of concepts (co-occurring concepts)
- Categorization: Relationships to semantic types from the Semantic Network


Semantic Types
Anatomical Structure
Fully Formed Anatomical Structure
Embryonic Structure
Disease or Syndrome
Body Part, Organ or Organ Component
Pharmacologic Substance
Population Group
Semantic Network
Metathesaurus
Medias-tinum
4
Saccular Viscus
Angina Pectoris
97
Esophagus
12
Heart
Cardiotonic Agents
225
Left Phrenic Nerve
Tissue Donors
22
Heart Valves
9
Fetal Heart
31
Concepts

Metathesaurus files Relationships

- Asserted relationships — MRREL 
- Statistical relationships — MRCOC 

- Categorization — MRSTY 

Outline of Tutorial

- Why customize? — Betsy Humphreys
- Metathesaurus basics — Olivier Bodenreider
- How to customize?
  - Customize sources (MetamorphoSys) — Laura Roth
  - Customize strings — Olivier Bodenreider
  - Customize synonyms
  - Customize relationships
  - Customize concept spaces
- Adding “local” terminology — Bill Hole

What is MetamorphoSys?

- A tool distributed for use with the UMLS Knowledge Sources
  - Already present in UMLS distribution in META/METAMSYS directory
- Multi-platform Java software
- Creates a customized version of the Metathesaurus
- An updated version has been created for 2002 release
  - Simpler to use with more features

Why use MetamorphoSys?

- Exclude vocabularies as required by the UMLS License Agreement
  - Default action is to select only vocabularies that have no additional restrictions (category zero)
- Remove terminology that may not fit a particular view or application
  - LOINC terms may be removed for Natural Language Processing
- Alter default “preferred name” precedence and control suppressibility of source term types

Why use MetamorphoSys?

- Remove relationships
  - e.g. Relationships from CCPSS not needed in application due to nature of rels and # that exist
- Currently, using MetamorphoSys, users cannot remove relationships from a particular source without removing all other data
  - In example above, to remove CCPSS relationships would remove all CCPSS data using the interface
  - Future MetamorphoSys enhancements may allow for removal of only relationships

How does MetamorphoSys Work?

- What it does: removes all information from MR* files that is supplied by the excluded vocabularies
  - This includes strings, relationships, attributes, mappings, etc.
- What results: A full Metathesaurus, including all the MR* files, containing information that matches what the user requested

How to Use MetamorphoSys

- Machine requirements
- Graphical User Interface
- Customizing with the interface

Machine Requirements

- A minimum of 256 MB of physical memory, as well as 8 GB recommended free disk space
  - Full UMLS distribution needs to be present
  - MetamorphoSys needs to be in the same directory as the data
- Can run on all Java platforms

Graphical User Interface

- Uses a Java graphical user interface
- Started by the MetamorphoSys program once UMLS distribution has been unpacked
  - Found in the /META/METAMSYS directory
  - MetamorphoSys.sh starts the program in the UNIX environment
  - MetamorphoSys.bat starts the program in the Windows environment

Graphical User Interface

- Simple to use
  - Allows users to make changes and save the changes for later use without having to edit a config file
- Composed of 4 Tabs
- Default is a Metathesaurus with just category zero vocabularies
  - Restriction levels are listed in License Agreement and are also listed in the interface under the Sources tab

Graphical User Interface components

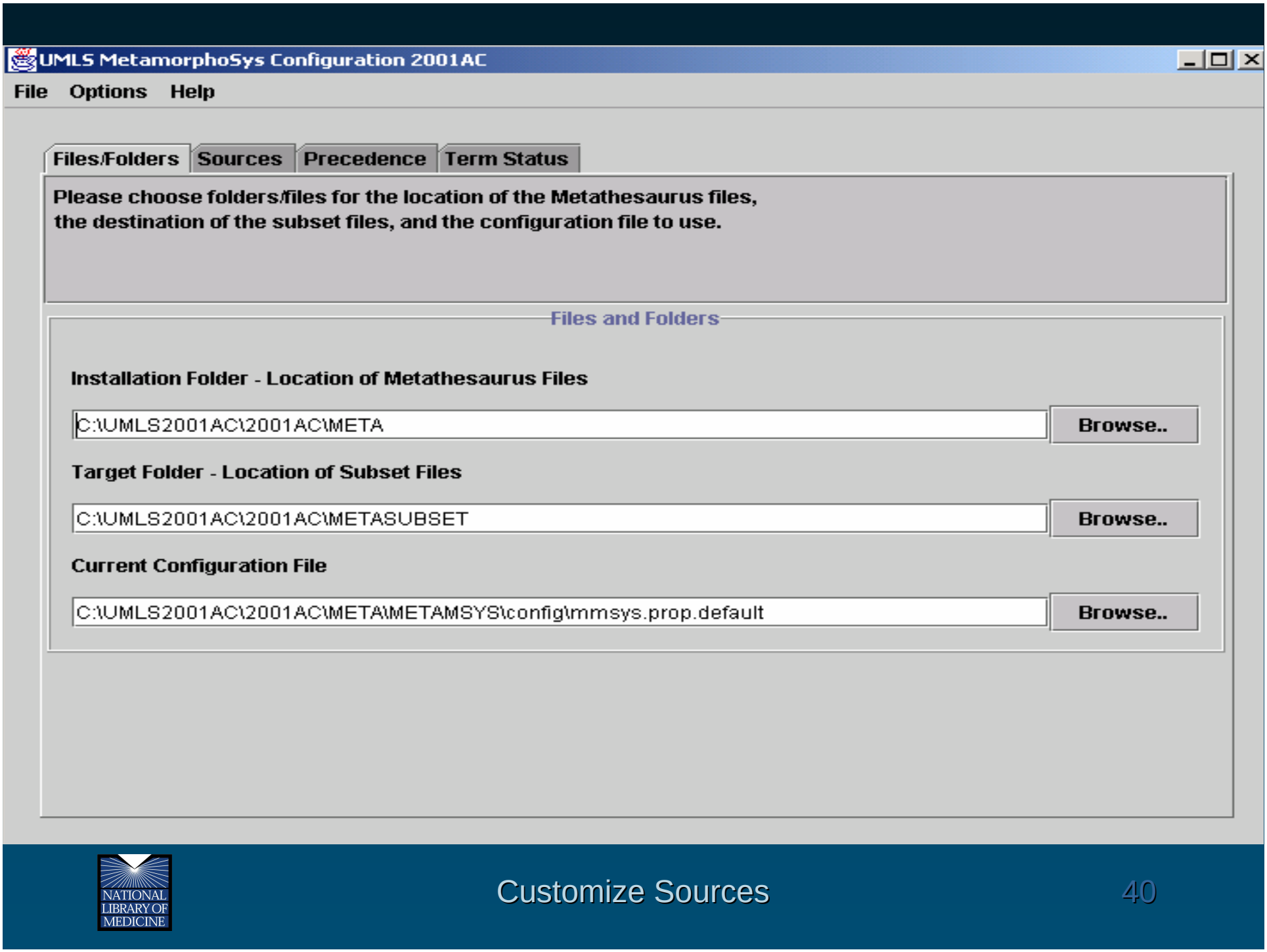
- Four tabs and an Options menu are present in the interface
  - Files/Folders
  - Sources
  - Precedence
  - Term Status
  - Options menu
    - Reset default settings
    - Advanced Options menu
    - Edit precedence

Files/ Folders tab

- MetamorphoSys is version aware
  - Links to Metathesaurus version it should be run against
  - On the top bar of the interface, the Meta version that should be used is listed
  - If a user tries to run against another version, a warning message appears

Files/ Folders tab

- Indicate where UMLS distribution is located
- Indicate where the customized Metathesaurus should go
- Indicate which config file should be used (default is the config file that came with MetamorphoSys but users can select their own)
- Default directories are provided but users can change if needed


UMLS MetamorphoSys Configuration 2001AC
File
Options
Help
Files/Folders
Sources
Precedence
Term Status
Please choose folders/files for the location of the Metathesaurus files,
the destination of the subset files, and the configuration file to use.
Files and Folders
Installation Folder - Location of Metathesaurus Files
C:\UMLS2001AC\2001AC\META
Browse..
Target Folder - Location of Subset Files
C:\UMLS2001AC\2001AC\METASUBSET
Browse..
Current Configuration File
C:\UMLS2001AC\2001AC\META\METAMSYS\config\mmsys.prop.default
Browse..

Sources Tab

- Sources are listed alphabetically
  - Includes full source name, abbreviation, Source Family Name and restriction level
  - Can be sorted on any of these fields
- Sources highlighted are the ones to be excluded
- Can change to include or exclude any vocabulary
  - The <ctrl> key needs to be held down to select or deselect new sources
- Options menu allows default values to be reset

Sources Tab Source Family Value

- Sources are now assigned a Source Family Value
  - All related sources are given the same Family Value
  - This allows sources to be grouped together that are covered under the same licensing agreements
  - For example: WHOART and all its foreign language versions (they all have a source family value of WHO)

Sources Tab Source Family Value

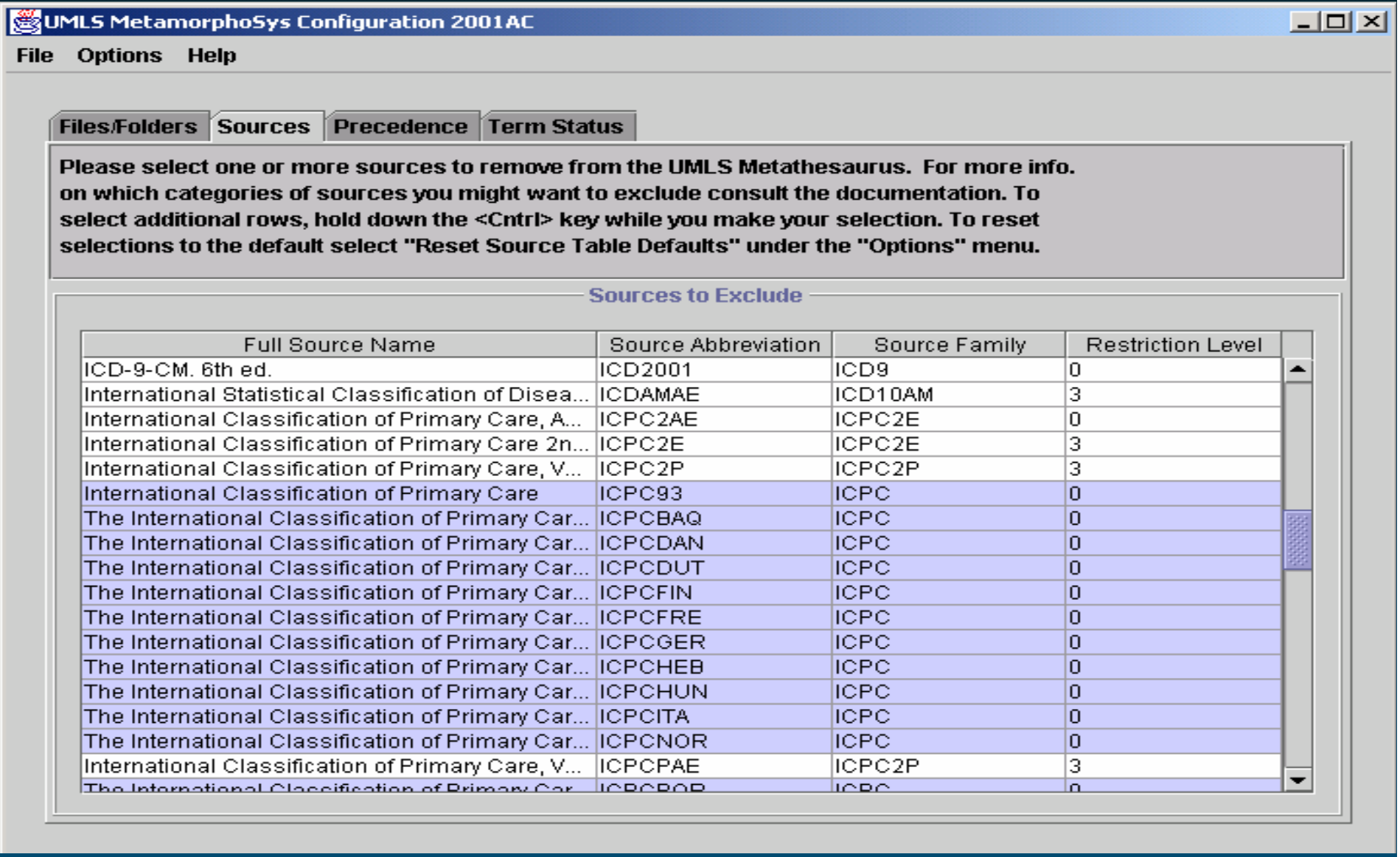
- When you click on one member of a source family, another window will appear verifying that all members of that family will be removed
  - Default is that all family members are removed but this can be changed
- Under Advanced Options, user can deactivate enforcement of family selection
  - Can also select auto-enforcement which will not give the user a chance to deselect any source family members

Sources Tab Dependent Source Value

- Sources can also have a Dependent Source value
  - Sometimes sources are related in a way similar to source families but do not properly belong in the same family. These are grouped together so they can be removed together if needed
    - e.g. CPT (family=CPT) and HCPT (family=HCPCS)
  - Advanced Options allows users to create their own dependent source relationships

Sources Tab Dependent Source Value

- When you click on one member of a dependent source, another window will appear verifying that all members of that dependent source will be removed
  - Default is that all members are removed but this can be changed
- Under Advanced Options, user can deactivate enforcement of dependent source selection
  - Can also select auto-enforcement which will not give the user a chance to deselect any dependent source members

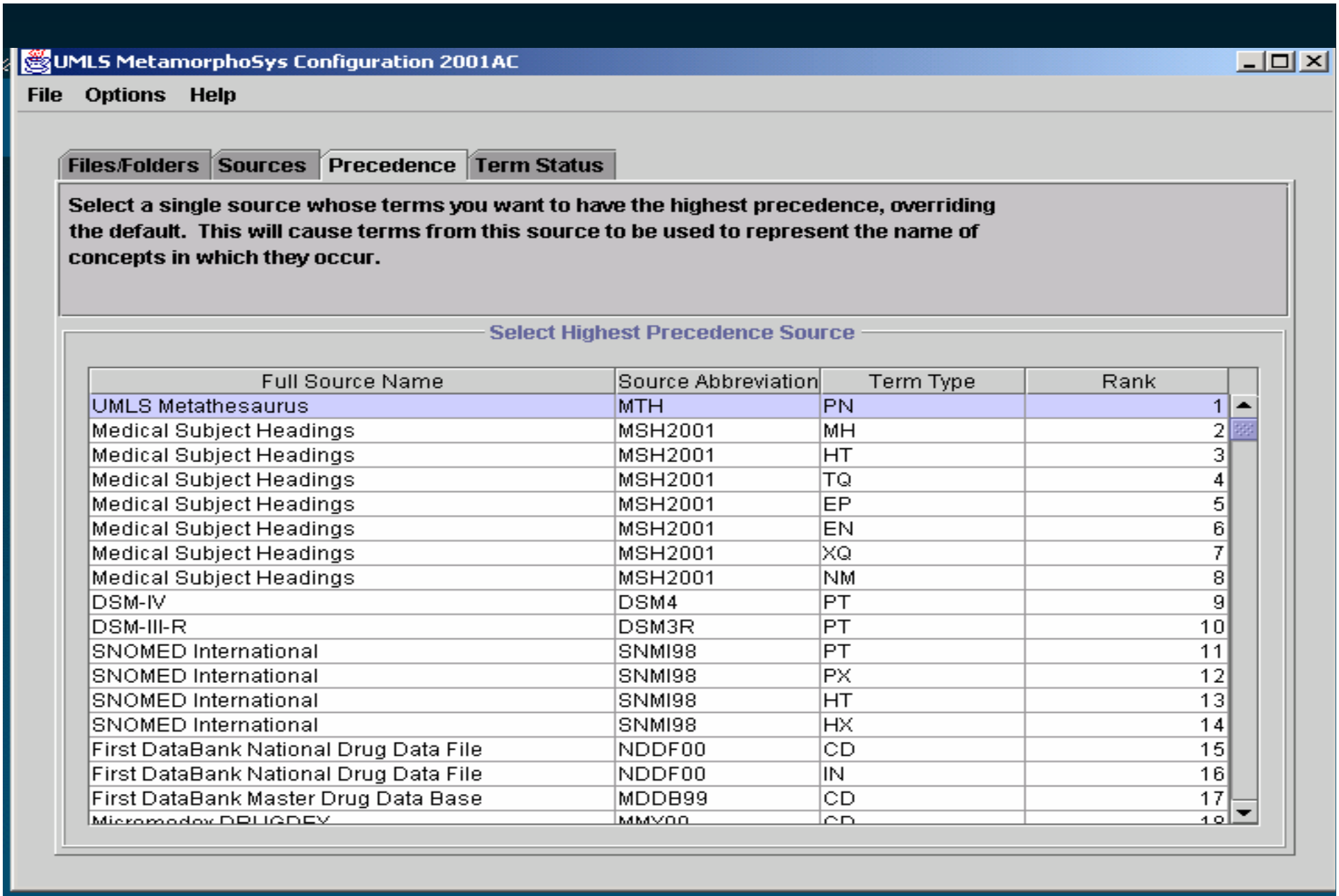

UMLS MetamorphoSys Configuration 2001AC
File Options Help
Files/Folders Sources Precedence Term Status
Please select one or more sources to remove from the UMLS Metathesaurus. For more info. on which categories of sources you might want to exclude consult the documentation. To select additional rows, hold down the <Cntrl> key while you make your selection. To reset selections to the default select "Reset Source Table Defaults" under the "Options" menu.
Sources to Exclude
Full Source Name | Source Abbreviation | Source Family | Restriction Level
ICD-9-CM. 6th ed. | ICD2001 | ICD9 | 0
International Statistical Classification of Disea... | ICDAMAE | ICD10AM | 3
International Classification of Primary Care, A... | ICPC2AE | ICPC2E | 0
International Classification of Primary Care 2n... | ICPC2E | ICPC2E | 3
International Classification of Primary Care, V... | ICPC2P | ICPC2P | 3
International Classification of Primary Care | ICPC93 | ICPC | 0
The International Classification of Primary Car... | ICPCBAQ | ICPC | 0
The International Classification of Primary Car... | ICPCDAN | ICPC | 0
The International Classification of Primary Car... | ICPCDUT | ICPC | 0
The International Classification of Primary Car... | ICPCFIN | ICPC | 0
The International Classification of Primary Car... | ICPCFRE | ICPC | 0
The International Classification of Primary Car... | ICPCGER | ICPC | 0
The International Classification of Primary Car... | ICPCHEB | ICPC | 0
The International Classification of Primary Car... | ICPCHUN | ICPC | 0
The International Classification of Primary Car... | ICPCITA | ICPC | 0
The International Classification of Primary Car... | ICPCNOR | ICPC | 0
International Classification of Primary Care, V... | ICPCPAE | ICPC2P | 3

Precedence Tab

- MTH source is the default highest precedence source
- Sources are arranged by their rank with highest rank first
- Fields include full source name, source abbreviation, term type and rank
  - Table can be sorted on any of these fields
- Highlighting a source will select it as the highest precedence
  - Only one source can be chosen at a time

Precedence Tab

- Options menu allows user to Edit Precedence
  - This opens a new window listing all the sources and term types in ranked order with MTH/PN as the highest
  - Users cut and paste the source-term types into whatever order they want
  - This new order can be saved by users in their own config file

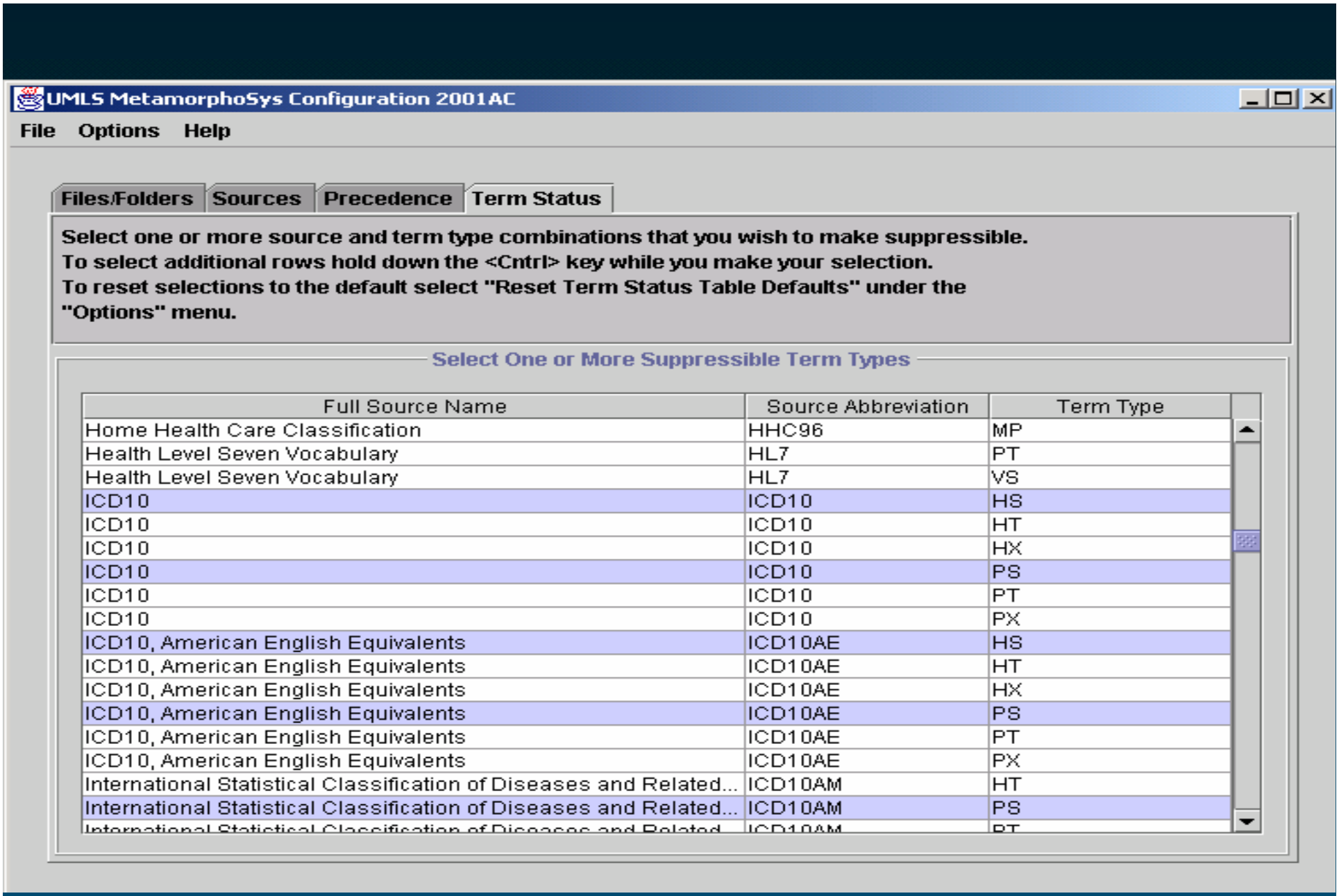

UMLS MetamorphoSys Configuration 2001AC
File
Options
Help
Files/Folders
Sources
Precedence
Term Status
Select a single source whose terms you want to have the highest precedence, overriding the default. This will cause terms from this source to be used to represent the name of concepts in which they occur.
Select Highest Precedence Source
Full Source Name
Source Abbreviation
Term Type
Rank
UMLS Metathesaurus MTH PN 1
Medical Subject Headings MSH2001 MH 2
Medical Subject Headings MSH2001 HT 3
Medical Subject Headings MSH2001 TQ 4
Medical Subject Headings MSH2001 EP 5
Medical Subject Headings MSH2001 EN 6
Medical Subject Headings MSH2001 XQ 7
Medical Subject Headings MSH2001 NM 8
DSM-IV DSM4 PT 9
DSM-III-R DSM3R PT 10
SNOMED International SNMI98 PT 11
SNOMED International SNMI98 PX 12
SNOMED International SNMI98 HT 13
SNOMED International SNMI98 HX 14
First DataBank National Drug Data File NDDF00 CD 15
First DataBank National Drug Data File NDDF00 IN 16
First DataBank Master Drug Data Base MDDB99 CD 17

Term Status Tab

- Used to add suppressibility
- All source-term type combinations that are suppressible are highlighted
- Cannot change term types that are already suppressible to non-suppressible
- New combinations can be highlighted to make suppressible

Term Status Tab

- Under Advanced Options, a user can now choose to remove all suppressible data from the subsetted Metathesaurus being created
- If not removed, the data is just marked as suppressible with a little “s”

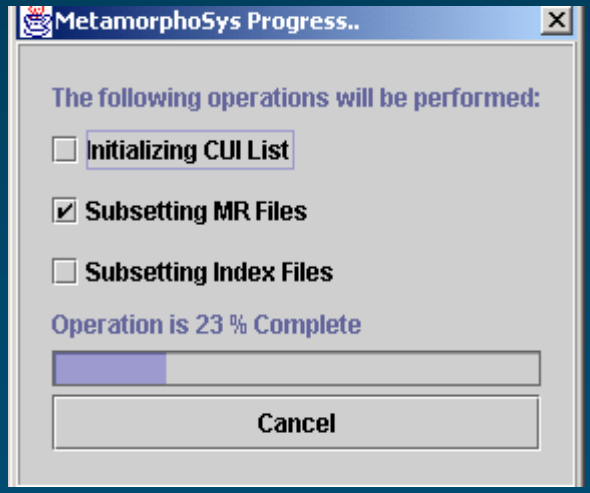

UMLS MetamorphoSys Configuration 2001AC
File
Options
Help
Files/Folders
Sources
Precedence
Term Status
Select one or more source and term type combinations that you wish to make suppressible.
To select additional rows hold down the <Cntrl> key while you make your selection.
To reset selections to the default select "Reset Term Status Table Defaults" under the "Options" menu.
Select One or More Suppressible Term Types
Full Source Name
Source Abbreviation
Term Type
Home Health Care Classification HHC96 MP
Health Level Seven Vocabulary HL7 PT
Health Level Seven Vocabulary HL7 VS
ICD10 ICD10 HS
ICD10 ICD10 HT
ICD10 ICD10 HX
ICD10 ICD10 PS
ICD10 ICD10 PT
ICD10 ICD10 PX
ICD10, American English Equivalents ICD10AE HS
ICD10, American English Equivalents ICD10AE HT
ICD10, American English Equivalents ICD10AE HX
ICD10, American English Equivalents ICD10AE PS
ICD10, American English Equivalents ICD10AE PT
ICD10, American English Equivalents ICD10AE PX
International Statistical Classification of Diseases and Related... ICD10AM HT
International Statistical Classification of Diseases and Related... ICD10AM PS

Running MetamorphoSys

- Once configuration is defined, a simple file selection starts subsetting
  - Under File Menu – Begin MetamorphoSys
- Before subsetting begins, user is asked if they want the current config file (with all changes) to be saved
  - This is how a user can save changes for future runs of MetamorphoSys

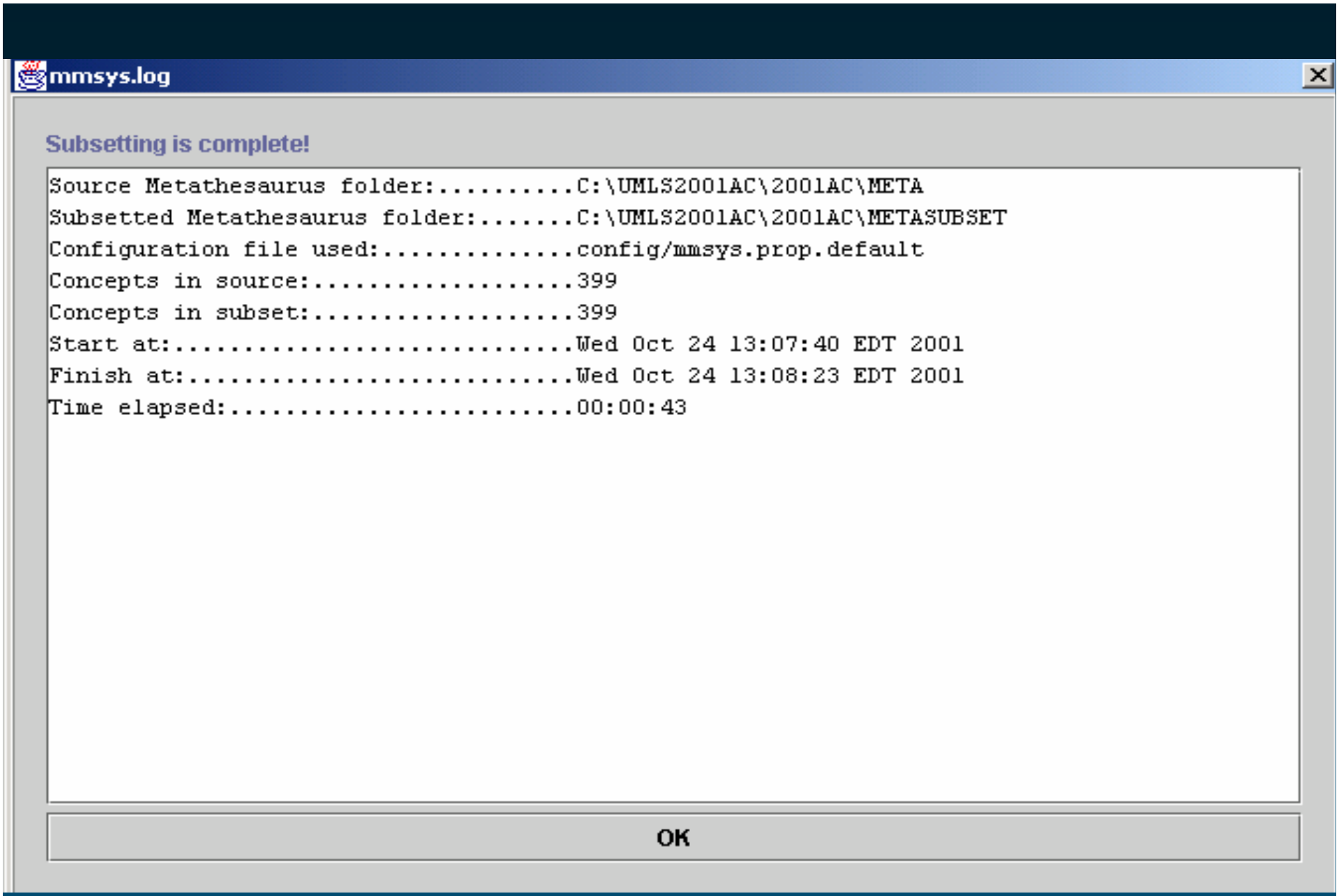
Progress Monitor

- Once subsetting begins, a progress monitor tracks process
  - Tracks progress through three major steps
  - Screen disappears only when subsetting is complete
  - “Cancel” ends the subsetting process



Log File

- After completion, a log file screen appears to indicate the process is complete and will report any errors
  - Log lists data files used, where the subsetted Metathesaurus is, name of configuration file used, number of concepts in subsetted files, time elapsed
  - Found in subset directory


mmsys.log
Subsetting is complete!
Source Metathesaurus folder:.....C:\UMLS2001AC\2001AC\META
Subsetted Metathesaurus folder:.....C:\UMLS2001AC\2001AC\METASUBSET
Configuration file used:.....config/mmsys.prop.default
Concepts in source:.....399
Concepts in subset:.....399
Start at:.....Wed Oct 24 13:07:40 EDT 2001
Finish at:.....Wed Oct 24 13:08:23 EDT 2001
Time elapsed:.....00:00:43
OK

For More MetamorphoSys Information

- See README Appendix B in the tutorial handout
- Go to http://umlsinfo.nlm.nih.gov and click on the UMLS Tools section
- Read Section 2.8 of the UMLS Documentation

Outline of Tutorial

- Why customize? — Betsy Humphreys
- Metathesaurus basics — Olivier Bodenreider
- How to customize?
  - Customize sources — Suresh Srinivasan
  - Customize strings — Olivier Bodenreider
  - Customize synonyms
  - Customize relationships
  - Customize concept spaces
- Adding “local” terminology — Bill Hole

MetamorphoSys Details

- MetamorphoSys output for:
  - Source exclusion
  - Altering precedence
  - Adding to suppressibility
- MetamorphoSys Configuration
- Looking ahead

Metathesaurus Data for **C0001403 (“Addison’s Disease ”)**

MRCON, MRSO Data for C0001403

MRCON

```
C0001403|ENG|P|L0001403|PF|S0010794|Addison's Disease|0|
C0001403|ENG|P|L0001403|VC|S0352253|ADDISON'S DISEASE|0|
C0001403|ENG|P|L0001403|VO|S0033587|Disease, Addison|0|
C0001403|ENG|P|L0001403|VO|S0469271|Addison's disease, NOS|3|
C0001403|ENG|S|L0367999|PF|S0469267|Addison melanoderma|3|
C0001403|ENG|S|L0373744|PF|S0471237|Asthenia pigmentosa|3|
```

MRSO

```
C0001403|L0001403|S0010794|MSH2001|MH|D000224|0|
C0001403|L0001403|S0352253|CST95|GT|ADREN INSUFFIC|0|
C0001403|L0001403|S0352253|WHO97|IT|0410|2|
C0001403|L0001403|S0033587|MSH2001|PM|D000224|0|
C0001403|L0001403|S0469271|SNMI98|PT|DB-70620|3|
C0001403|L0367999|S0469267|SNMI98|SY|DB-70620|3|
C0001403|L0373744|S0471237|SNMI98|SY|DB-70620|3|
```

MRCON, MRSO Data for C0001403

MRCON

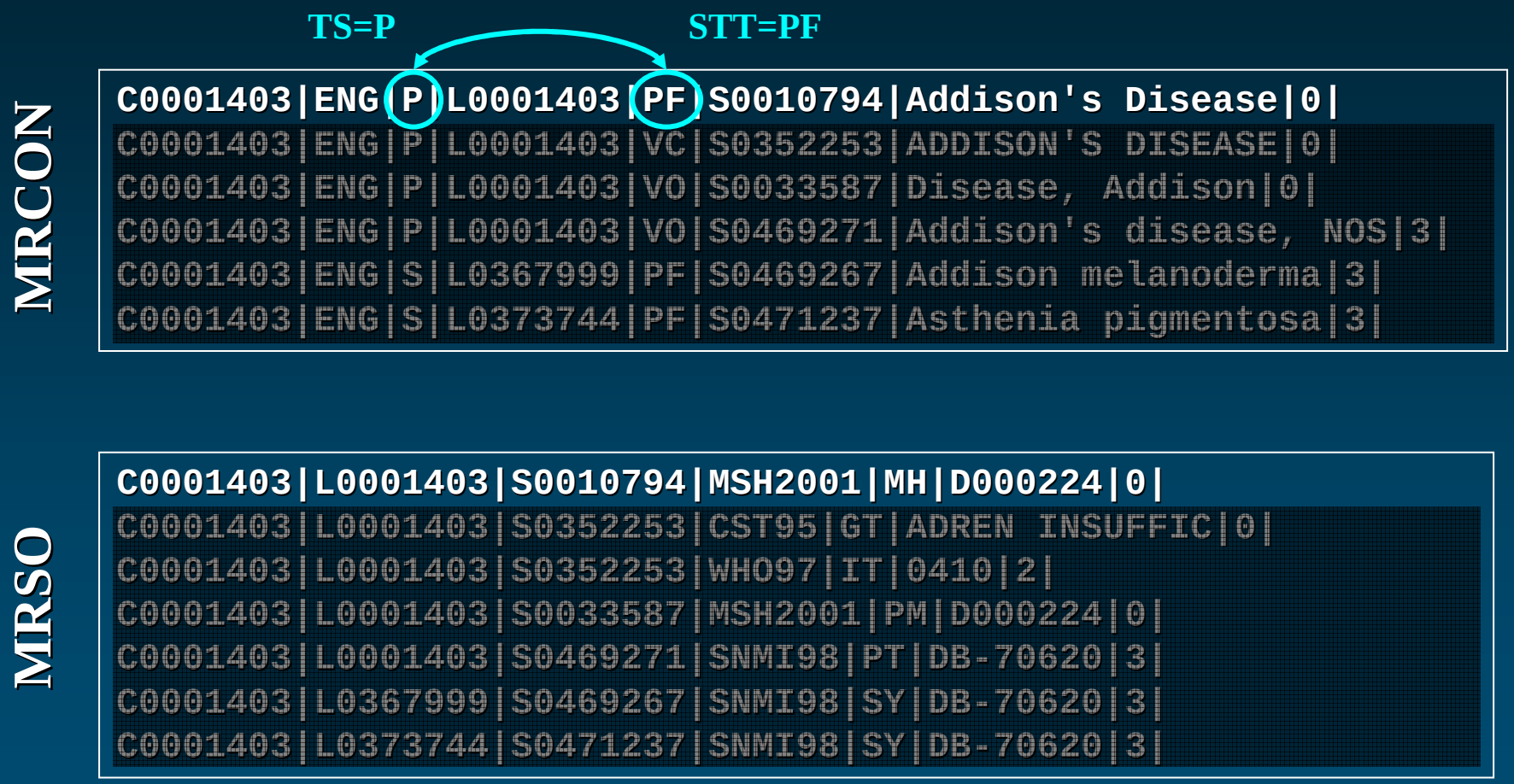
```
C0001403|ENG|P|L0001403|PF|S0010794|Addison's Disease|0|
C0001403|ENG|P|L0001403|VC|S0352253|ADDISON'S DISEASE|0|
C0001403|ENG|P|L0001403|VO|S0033587|Disease, Addison|0|
C0001403|ENG|P|L0001403|VO|S0469271|Addison's disease, NOS|3|
C0001403|ENG|S|L0367999|PF|S0469267|Addison melanoderma|3|
C0001403|ENG|S|L0373744|PF|S0471237|Asthenia pigmentosa|3|
```

MRSO

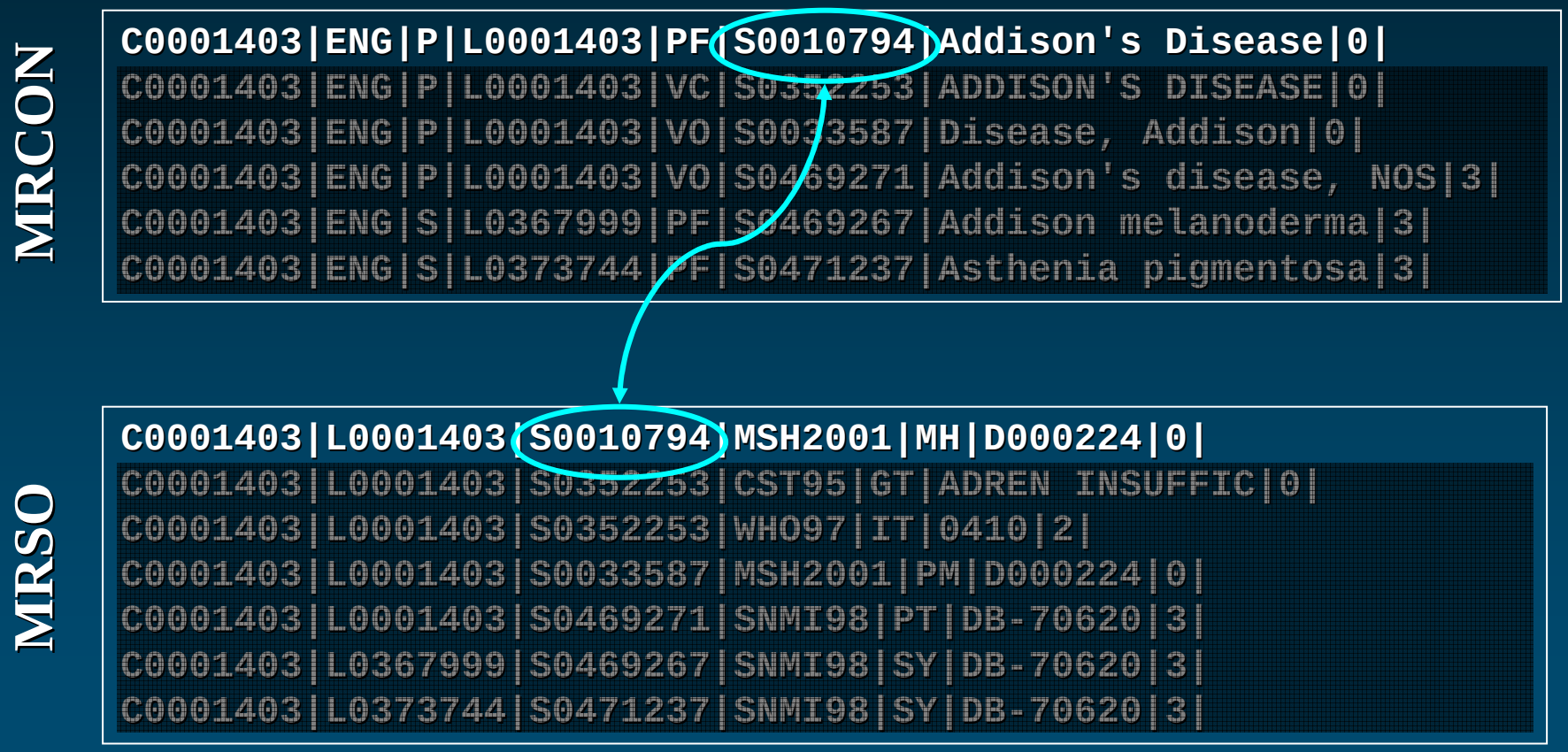
```
C0001403|L0001403|S0010794|MSH2001|MH|D000224|0|
C0001403|L0001403|S0352253|CST95|GT|ADREN INSUFFIC|0|
C0001403|L0001403|S0352253|WHO97|IT|0410|2|
C0001403|L0001403|S0033587|MSH2001|PM|D000224|0|
C0001403|L0001403|S0469271|SNMI98|PT|DB-70620|3|
C0001403|L0367999|S0469267|SNMI98|SY|DB-70620|3|
C0001403|L0373744|S0471237|SNMI98|SY|DB-70620|3|
```

MRCON, MRSO Data for C0001403

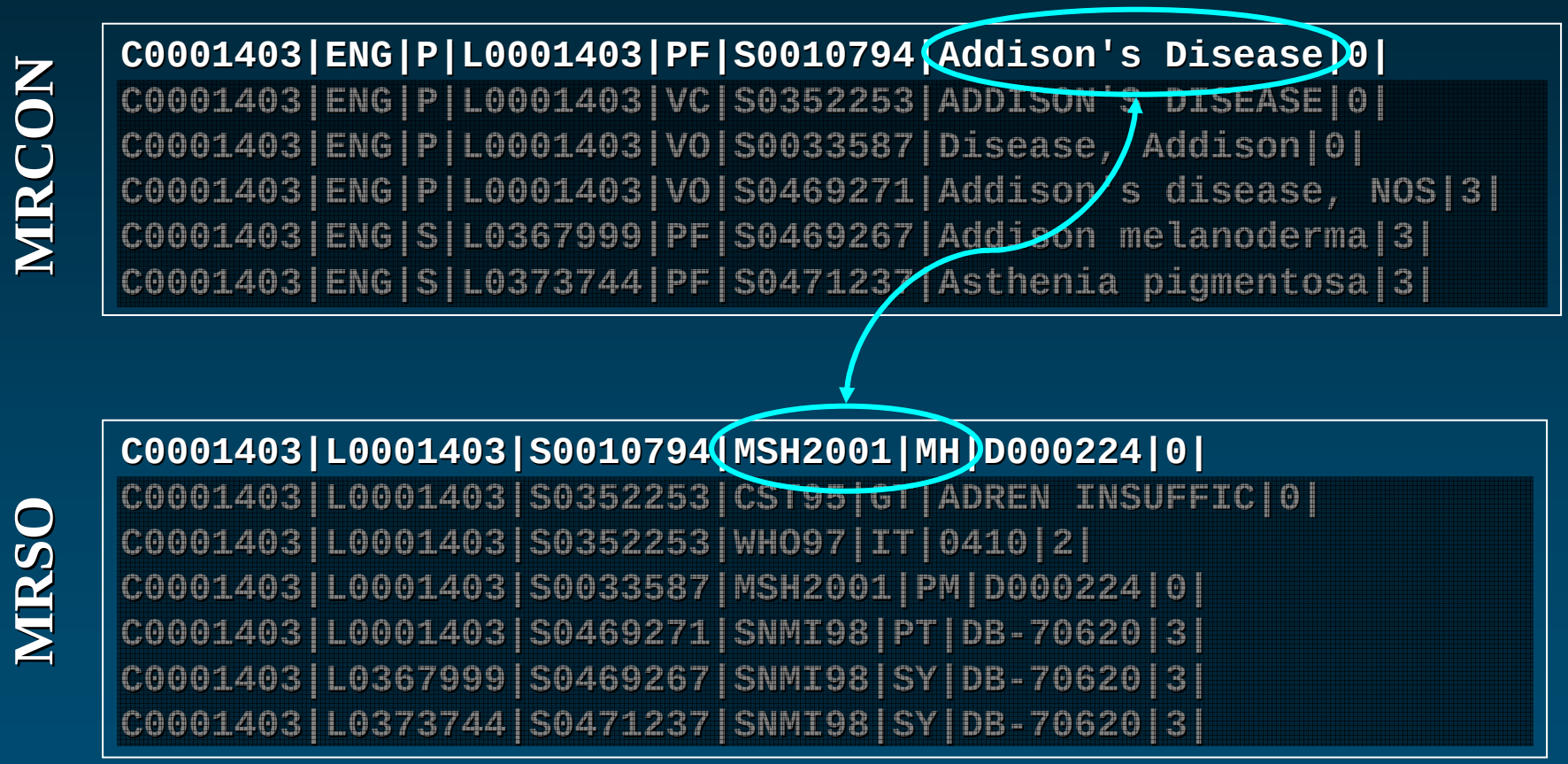
MRCON, MRSO Data for C0001403



MRCON, MRSO Data for C0001403



MRCON, MRSO Data for C0001403



MRCON, MRSO Data for C0001403

MRCON

```
C0001403|ENG|P|L0001403|PF|S0010794|Addison's Disease|0|
C0001403|ENG|P|L0001403|VC|S0352253|ADDISON'S DISEASE|0|
C0001403|ENG|P|L0001403|VO|S0033587|Disease, Addison|0|
C0001403|ENG|P|L0001403|VO|S0469271|Addison's disease, NOS|3|
C0001403|ENG|S|L0367999|PF|S0469267|Addison melanoderma|3|
C0001403|ENG|S|L0373744|PF|S0471237|Asthenia pigmentosa|3|
```

MRSO

```
C0001403|L0001403|S0010794|MSH2001|MH|D000224|0|
C0001403|L0001403|S0352253|CST95|GT|ADREN INSUFFIC|0|
C0001403|L0001403|S0352253|WHO97|IT|0410|2|
C0001403|L0001403|S0033587|MSH2001|PM|D000224|0|
C0001403|L0001403|S0469271|SNMI98|PT|DB-70620|3|
C0001403|L0367999|S0469267|SNMI98|SY|DB-70620|3|
C0001403|L0373744|S0471237|SNMI98|SY|DB-70620|3|
```

MRCON, MRSO Data for C0001403

MRREL, MRSAT Data for C0001403

MRREL

```
C0001403|CHD|C0546992||RCD99|RCD99||
C0001403|PAR|C0001621||PSY2001|PSY2001||
C0001403|PAR|C0004364|inverse_isa|MSH2001|MSH2001||
C0001403|RB|C0001621||MTH|MTH||
C0001403|RB|C0004364||CSP2001|CSP2001||
C0001403|RN|C0518933||MTH|MTH||
C0001403|RO|C0085860||MTH|MTH||
C0001403|RO|C0546992|associated_with|SNMI98|SNMI98||
```

MRSAT

```
C0001403|L0001403|S0010794|D000224|MN|MSH2001|C20.111.163|
C0001403|L0001403|S0010794|D000224|MUI|MSH2001|M0000346|
C0001403|L0001403|S0469271|DB-70620|SIC|SNMI98|255.4|
C0001403|L0001403|S1619433|10013096|MPC|MDR33|10001390|
```

MRREL, MRSAT Data for C0001403

MRREL

```
C0001403|CHD|C0546992||RCD99|RCD99||
C0001403|PAR|C0001621||PSY2001|PSY2001||
C0001403|PAR|C0004364|inverse_isa|MSH2001|MSH2001||
C0001403|RB|C0001621||MTH|MTH||
C0001403|RB|C0004364||CSP2001|CSP2001||
C0001403|RN|C0518933||MTH|MTH||
C0001403|RO|C0085860||MTH|MTH||
C0001403|RO|C0546992|associated_with|SNMI98|SNMI98||
```

Addison's Disease

<has child>

Tuberculous Addison's disease

MRSAT

```
C0001403|L0001403|S0010794|D000224|MN|MSH2001|C20.111.163|
C0001403|L0001403|S0010794|D000224|MUI|MSH2001|M0000346|
C0001403|L0001403|S0469271|DB-70620|SIC|SNMI98|255.4|
C0001403|L0001403|S1619433|10013096|MPC|MDR33|10001390|
```

MRREL, MRSAT Data for C0001403

MRREL

```
C0001403|CHD|C0546992||RCD99|RCD99||
C0001403|PAR|C0001621||PSY2001|PSY2001||
C0001403|PAR|C0004364|inverse_isa|MSH2001|MSH2001||
C0001403|RB|C0001621||MTH|MTH||
C0001403|RB|C0004364||CSP2001|CSP2001||
C0001403|RN|C0518933||MTH|MTH||
C0001403|RO|C0085860||MTH|MTH||
C0001403|RO|C0546992|associated_with|SNMI98|SNMI98||
```

Context Relationships from Sources

MRSAT

```
C0001403|L0001403|S001079                SH2001|C20.111.163|
C0001403|L0001403|S0010794|D000224|MUI|MSH2001|M0000346|
C0001403|L0001403|S0469271|DB-70620|SIC|SNMI98|255.4|
C0001403|L0001403|S1619433|10013096|MPC|MDR33|10001390|
```

MRREL, MRSAT Data for C0001403

MRREL

```
C0001403|CHD|C0546992||RCD99|RCD99||
C0001403|PAR|C0001621||PSY2001|PSY2001||
C0001403|PAR|C0004364|inverse_isa|MSH2001|MSH2001||
C0001403|RB|C0001621||MTH|MTH||
C0001403|RB|C0004364||CSP2001|CSP2001||
C0001403|RN|C0518933||MTH|MTH||
C0001403|RO|C0085860||MTH|MTH||
C0001403|RO|C0546992|associated_with|SNMI98|SNMI98||
```

Other Relationships from Sources and MTH

MRSAT

```
C0001403|L0001403|S001079                SH2001|C20.111.163|
C0001403|L0001403|S001079                MSH2001|M0000346|
C0001403|L0001403|S046927[illegible]|SNMI98|255.4|
C0001403|L0001403|S1619433|10013096|MPC|MDR33|10001390|
```

MRREL, MRSAT Data for C0001403

MRREL

```
C0001403|CHD|C0546992||RCD99|RCD99||
C0001403|PAR|C0001621||PSY2001|PSY2001||
C0001403|PAR|C0004364|inverse_isa|MSH2001|MSH2001||
C0001403|RB|C0001621||MTH|MTH||
C0001403|RB|C0004364||CSP2001|CSP20
C0001403|RN|C0518933||MTH|MTH||
C0001403|RO|C0085860||MTH|MTH||
C0001403|RO|C0546992|associated_with|SNMI98|SNMI98||
```

Source Attributes

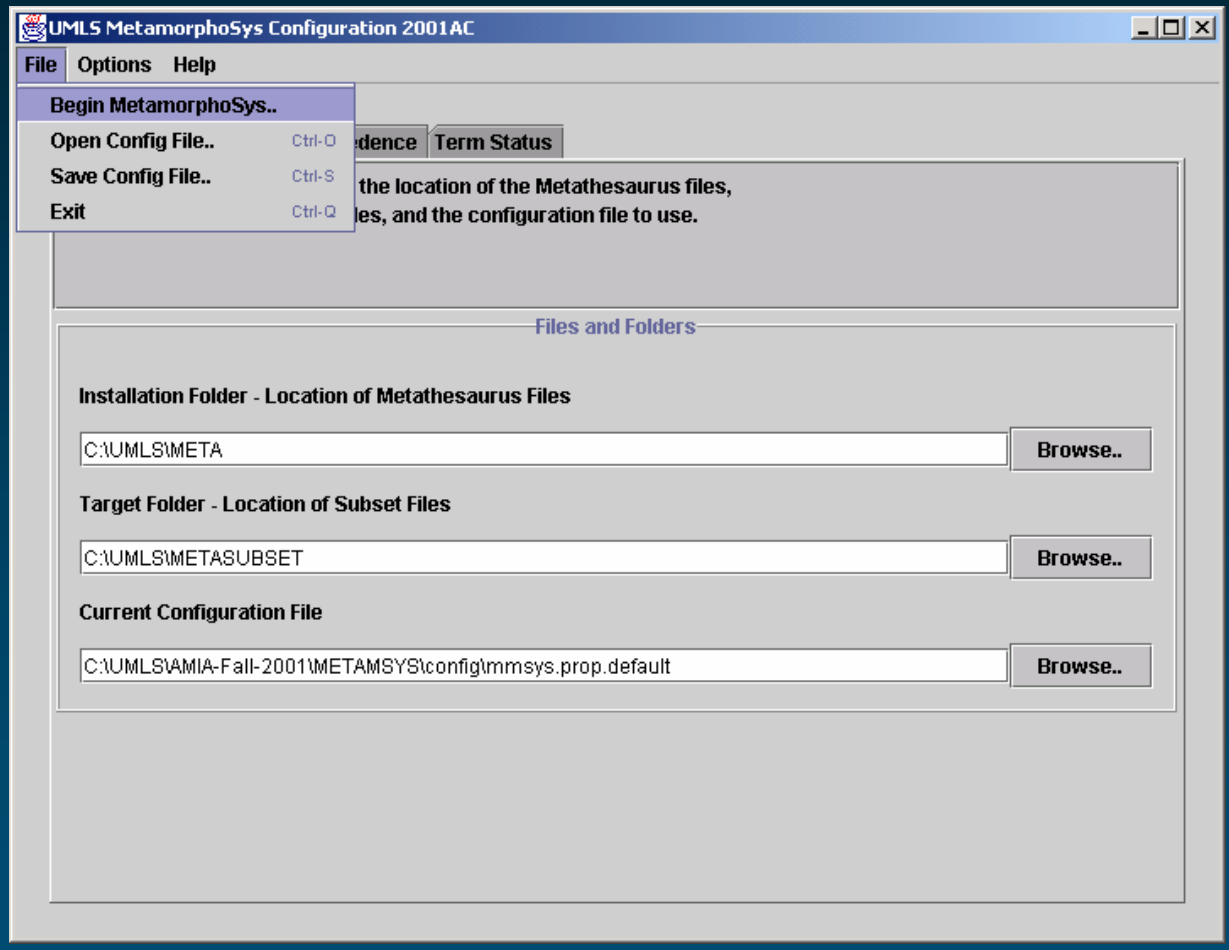
MRSAT

```
C0001403|L0001403|S0010794|D000224|MN|MSH2001|C20.111.163|
C0001403|L0001403|S0010794|D000224|MUI|MSH2001|M0000346|
C0001403|L0001403|S0469271|DB-70620|SIC|SNMI98|255.4|
C0001403|L0001403|S1619433|10013096|MPC|MDR33|10001390|
```

Default Subset using MetamorphoSys

- Removing all sources with a Source Restriction Level greater than 0
- Using default precedence ranking from MRRANK (highest precedence is MTH/PN, etc.)
- Default suppressibility and retaining suppressible rows in MRCON as TS=s

Default Subset



Default Subset: MRCON, MRSO

MRCON

```
C0001403|ENG|P|L0001403|PF|S0010794|Addison's Disease|0|
C0001403|ENG|P|L0001403|VC|S0352253|ADDISON'S DISEASE|0|
C0001403|ENG|P|L0001403|VO|S0033587|Disease, Addison|0|
C0001403|ENG|P|L0001403|VO|S0469271|Addison's disease, NOS|3|
C0001403|ENG|S|L0367999|PF|S0469267|Addison melanoderma|3|
C0001403|ENG|S|L0373744|PF|S0471237|Asthenia pigmentosa|3|
```

MRSO

```
C0001403|L0001403|S0010794|MSH2001|MH|D000224|0|
C0001403|L0001403|S0352253|CST95|GT|ADREN INSUFFIC|0|
C0001403|L0001403|S0352253|WHO97|IT|0410|2|
C0001403|L0001403|S0033587|MSH2001|PM|D000224|0|
C0001403|L0001403|S0469271|SNMI98|PT|DB-70620|3|
C0001403|L0367999|S0469267|SNMI98|SY|DB-70620|3|
C0001403|L0373744|S0471237|SNMI98|SY|DB-70620|3|
```

Rows excluded: MRCON, MRSO

MRCON

```
C0001403|ENG|P|L0001403|PF|S0010794|Addison's Disease|0|
C0001403|ENG|P|L0001403|VC|S0352253|ADDISON'S DISEASE|0|
C0001403|ENG|P|L0001403|VO|S0033587|Disease, Addison|0|
C0001403|ENG|P|L0001403|VO|S0469271|Addison's disease, NOS|3|
C0001403|ENG|S|L0367999|PF|S0469267|Addison melanoderma|3|
C0001403|ENG|S|L0373744|PF|S0471237|Asthenia pigmentosa|3|
```

Restricted Sources

MRSO

```
C0001403|L0001403|S0010794|MSH2001|MH|D000224|0|
C0001403|L0001403|S0352253|CST95|GT|ADREN INSUFFIC|0|
C0001403|L0001403|S0352253|WHO97|IT|0410|2|
C0001403|L0001403|S0033587|MSH2001|PM|D000224|0|
C0001403|L0001403|S0469271|SNMI98|PT|DB-70620|3|
C0001403|L0367999|S0469267|SNMI98|SY|DB-70620|3|
C0001403|L0373744|S0471237|SNMI98|SY|DB-70620|3|
```

Rows remaining: MRCON, MRSO

MRCON

```
C0001403|ENG|P|L0001403|PF|S0010794|Addison's Disease|0|
C0001403|ENG|P|L0001403|VC|S0352253|ADDISON'S DISEASE|0|
C0001403|ENG|P|L0001403|VO|S0033587|Disease, Addison|0|
C0001403|ENG|P|L0001403|VO|S0469271|Addison's disease, NOS|3|
C0001403|ENG|S|L0367999|PF|S0469267|Addison melanoderma|3|
C0001403|ENG|S|L0373744|PF|S0471237|Asthenia pigmentosa|3|
```

MRSO

```
C0001403|L0001403|S0010794|MSH2001|MH|D000224|0|
C0001403|L0001403|S0352253|CST95|GT|ADREN INSUFFIC|0|
C0001403|L0001403|S0352253|WHO97|IT|0410|2|
C0001403|L0001403|S0033587|MSH2001|PM|D000224|0|
C0001403|L0001403|S0469271|SNMI98|PT|DB-70620|3|
C0001403|L0367999|S0469267|SNMI98|SY|DB-70620|3|
C0001403|L0373744|S0471237|SNMI98|SY|DB-70620|3|
```

Preferred name remains unchanged

TS=P STT=PF

MRCON

```
C0001403|ENG|P|L0001403|PF|S0010794|Addison's Disease|0|
C0001403|ENG|P|L0001403|VC|S0352253|ADDISON'S DISEASE|0|
C0001403|ENG|P|L0001403|VO|S0033587|Disease, Addison|0|
C0001403|ENG|P|L0001403|VO|S0469271|Addison's disease, NOS|3|
C0001403|ENG|S|L0367999|PF|S0469267|Addison melanoderma|3|
C0001403|ENG|S|L0373744|PF|S0471237|Asthenia pigmentosa|3|
```

MRSO

```
C0001403|L0001403|S0010794|MSH2001|MH|D000224|0|
C0001403|L0001403|S0352253|CST95|GT|ADREN INSUFFIC|0|
C0001403|L0001403|S0352253|WHO97|IT|0410|2|
C0001403|L0001403|S0033587|MSH2001|PM|D000224|0|
C0001403|L0001403|S0469271|SNMI98|PT|DB-70620|3|
C0001403|L0367999|S0469267|SNMI98|SY|DB-70620|3|
C0001403|L0373744|S0471237|SNMI98|SY|DB-70620|3|
```

S0352253 Survives

MRCON

```
C0001403|ENG|P|L0001403|PF|S0010794|Addison's Disease|0|
C0001403|ENG|P|L0001403|VC|S0352253|ADDISON'S DISEASE|0|
C0001403|ENG|P|L0001403|VO|S0033587|Disease, Addison|0|
C0001403|ENG|P|L0001403|VO|S0469271|Addison's disease, NOS|3|
C0001403|ENG|S|L0367999|PF|S0469267|Addison melanoderma|3|
C0001403|ENG|S|L0373744|PF|S0471237|Asthenia pigmentosa|3|
```

MRSO

```
C0001403|L0001403|S0010794|MSH2001|MH|D000224|0|
C0001403|L0001403|S0352253|CST95|GT|ADREN INSUFFIC|0|
C0001403|L0001403|S0352253|WHO97|IT|0410|2|
C0001403|L0001403|S0033587|MSH2001|PM|D000224|0|
C0001403|L0001403|S0469271|SNMI98|PT|DB-70620|3|
C0001403|L0367999|S0469267|SNMI98|SY|DB-70620|3|
C0001403|L0373744|S0471237|SNMI98|SY|DB-70620|3|
```

Default subset: MRREL, MRSAT

MRREL

```
C0001403|CHD|C0546992||RCD99|RCD99||
C0001403|PAR|C0001621||PSY2001|PSY2001||
C0001403|PAR|C0004364|inverse_isa|MSH2001|MSH2001||
C0001403|RB|C0001621||MTH|MTH||
C0001403|RB|C0004364||CSP2001|CSP2001||
C0001403|RN|C0518933||MTH|MTH||
C0001403|RO|C0085860||MTH|MTH||
C0001403|RO|C0546992|associated_with|SNMI98|SNMI98||
```

MRSAT

```
C0001403|L0001403|S0010794|D000224|MN|MSH2001|C20.111.163|
C0001403|L0001403|S0010794|D000224|MUI|MSH2001|M0000346|
C0001403|L0001403|S0469271|DB-70620|SIC|SNMI98|255.4|
C0001403|L0001403|S1619433|10013096|MPC|MDR33|10001390|
```

Rows Excluded: MRREL, MRSAT

MRREL

```
C0001403|CHD|C0546992||RCD99|RCD99||
C0001403|PAR|C0001621||PSY2001|PSY2001||
C0001403|PAR|C0004364|inverse_isa|MSH2001|MSH2001||
C0001403|RB|C0001621||MTH|MTH||
C0001403|RB|C0004364||CSP2001|CSP2001||
C0001403|RN|C0518933||MTH|MTH||
C0001403|RO|C0085860||MTH|MTH||
C0001403|RO|C0546992|associated_with|SNMI98|SNMI98||
```

MRSAT

```
C0001403|L0001403|S0010794|D000224|MN|MSH2001|C20.111.163|
C0001403|L0001403|S0010794|D000224|MUI|MSH2001|M0000346|
C0001403|L0001403|S0469271|DB-70620|SIC|SNMI98|255.4|
C0001403|L0001403|S1619433|10013096|MPC|MDR33|10001390|
```

Rows Remaining: MRREL, MRSAT

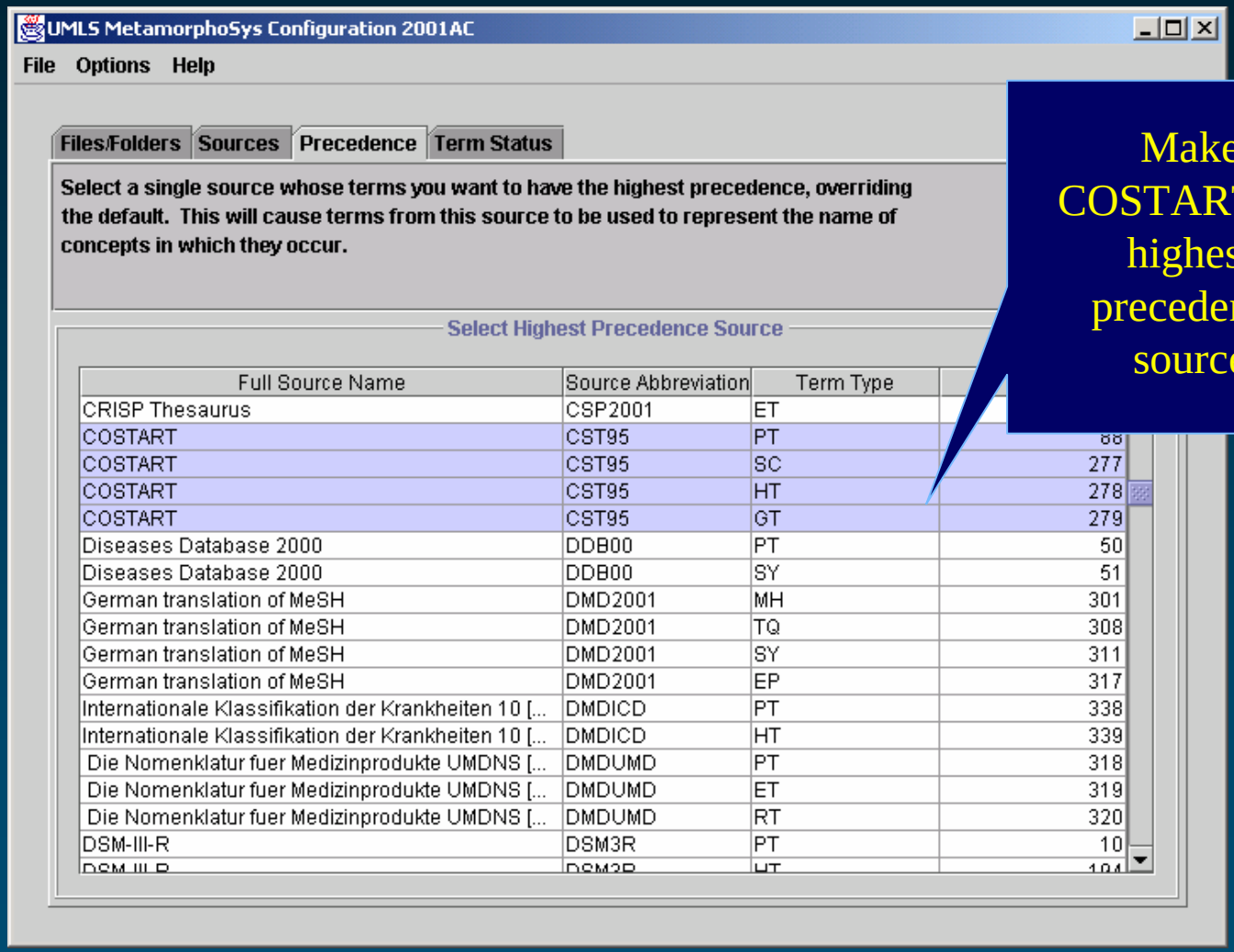
MRREL

```
C0001403|CHD|C0546992||RCD99|RCD99||
C0001403|PAR|C0001621||PSY2001|PSY2001||
C0001403|PAR|C0004364|inverse_isa|MSH2001|MSH2001||
C0001403|RB|C0001621||MTH|MTH||
C0001403|RB|C0004364||CSP2001|CSP2001||
C0001403|RN|C0518933||MTH|MTH||
C0001403|RO|C0085860||MTH|MTH||
C0001403|RO|C0546992|associated_with|SNMI98|SNMI98||
```

MRSAT

```
C0001403|L0001403|S0010794|D000224|MN|MSH2001|C20.111.163|
C0001403|L0001403|S0010794|D000224|MUI|MSH2001|M0000346|
C0001403|L0001403|S0469271|DB-70620|SIC|SNMI98|255.4|
C0001403|L0001403|S1619433|10013096|MPC|MDR33|10001390|
```

Changing Precedence



Preferred term changes from MeSH..

MRCON

```
C0001403|ENG|P|L0001403|PF|S0010794|Addison's Disease|0|
C0001403|ENG|P|L0001403|VC|S0352253|ADDISON'S DISEASE|0|
C0001403|ENG|P|L0001403|VO|S0033587|Disease, Addison|0|
C0001403|ENG|P|L0001403|VO|S0469271|Addison's disease, NOS|3|
C0001403|ENG|S|L0367999|PF|S0469267|Addison melanoderma|3|
C0001403|ENG|S|L0373744|PF|S0471237|Asthenia pigmentosa|3|
```

MRSO

```
C0001403|L0001403|S0010794|MSH2001|MH|D000224|0|
C0001403|L0001403|S0352253|CST95|GT|ADREN INSUFFIC|0|
C0001403|L0001403|S0352253|WHO97|IT|0410|2|
C0001403|L0001403|S0033587|MSH2001|PM|D000224|0|
C0001403|L0001403|S0469271|SNMI98|PT|DB-70620|3|
C0001403|L0367999|S0469267|SNMI98|SY|DB-70620|3|
C0001403|L0373744|S0471237|SNMI98|SY|DB-70620|3|
```

..to COSTART (CST95)

MRCON

```
C0001403|ENG|P|L0001403|PF|S0352253|ADDISON'S DISEASE|0|
C0001403|ENG|P|L0001403|VC|S0010794|Addison's Disease|0|
C0001403|ENG|P|L0001403|VO|S0033587|Disease, Addison|0|
C0001403|ENG|P|L0001403|VO|S0469271|Addison's disease, NOS|3|
C0001403|ENG|S|L0367999|PF|S0469267|Addison melanoderma|3|
C0001403|ENG|S|L0373744|PF|S0471237|Asthenia pigmentosa|3|
```

MRSO

```
C0001403|L0001403|S0010794|MSH2001|MH|D000224|0|
C0001403|L0001403|S0352253|CST95|GT|ADREN INSUFFIC|0|
C0001403|L0001403|S0352253|WHO97|IT|0410|2|
C0001403|L0001403|S0033587|MSH2001|PM|D000224|0|
C0001403|L0001403|S0469271|SNMI98|PT|DB-70620|3|
C0001403|L0367999|S0469267|SNMI98|SY|DB-70620|3|
C0001403|L0373744|S0471237|SNMI98|SY|DB-70620|3|
```

TS, STT and LRL get recomputed

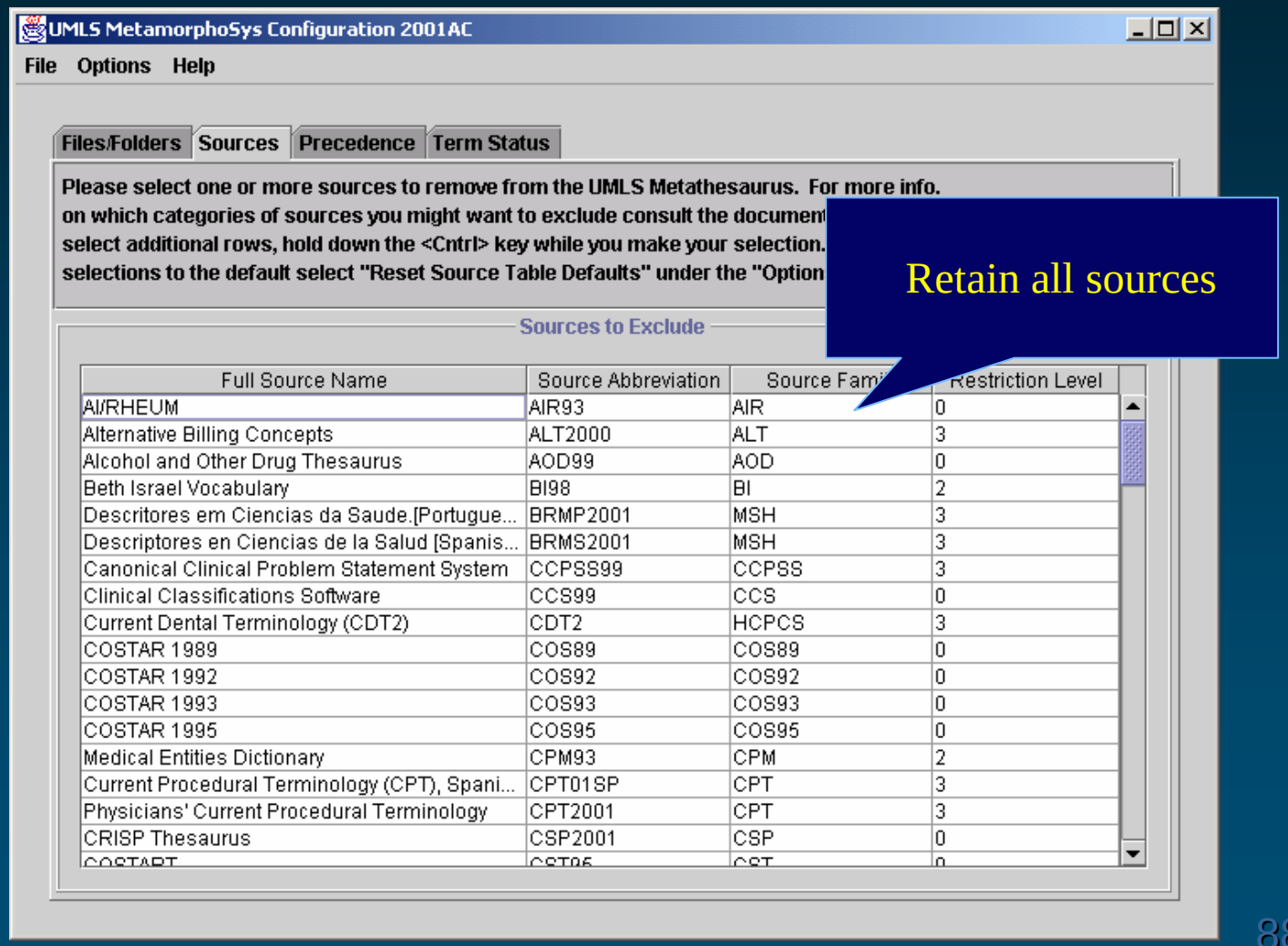
MRCON

```
C0001403|ENG|P|L0001403|PF|S0352253|ADDISON'S DISEASE|0|
C0001403|ENG|P|L0001403|VC|S0010794|Addison's Disease|0|
C0001403|ENG|P|L0001403|VO|S0033587|Disease, Addison|0|
C0001403|ENG|P|L0001403|VO|S0469271|Addison's disease, NOS|3|
C0001403|ENG|S|L0367999|PF|S0469267|Addison melanoderma|3|
C0001403|ENG|S|L0373744|PF|S0471237|Asthenia pigmentosa|3|
```

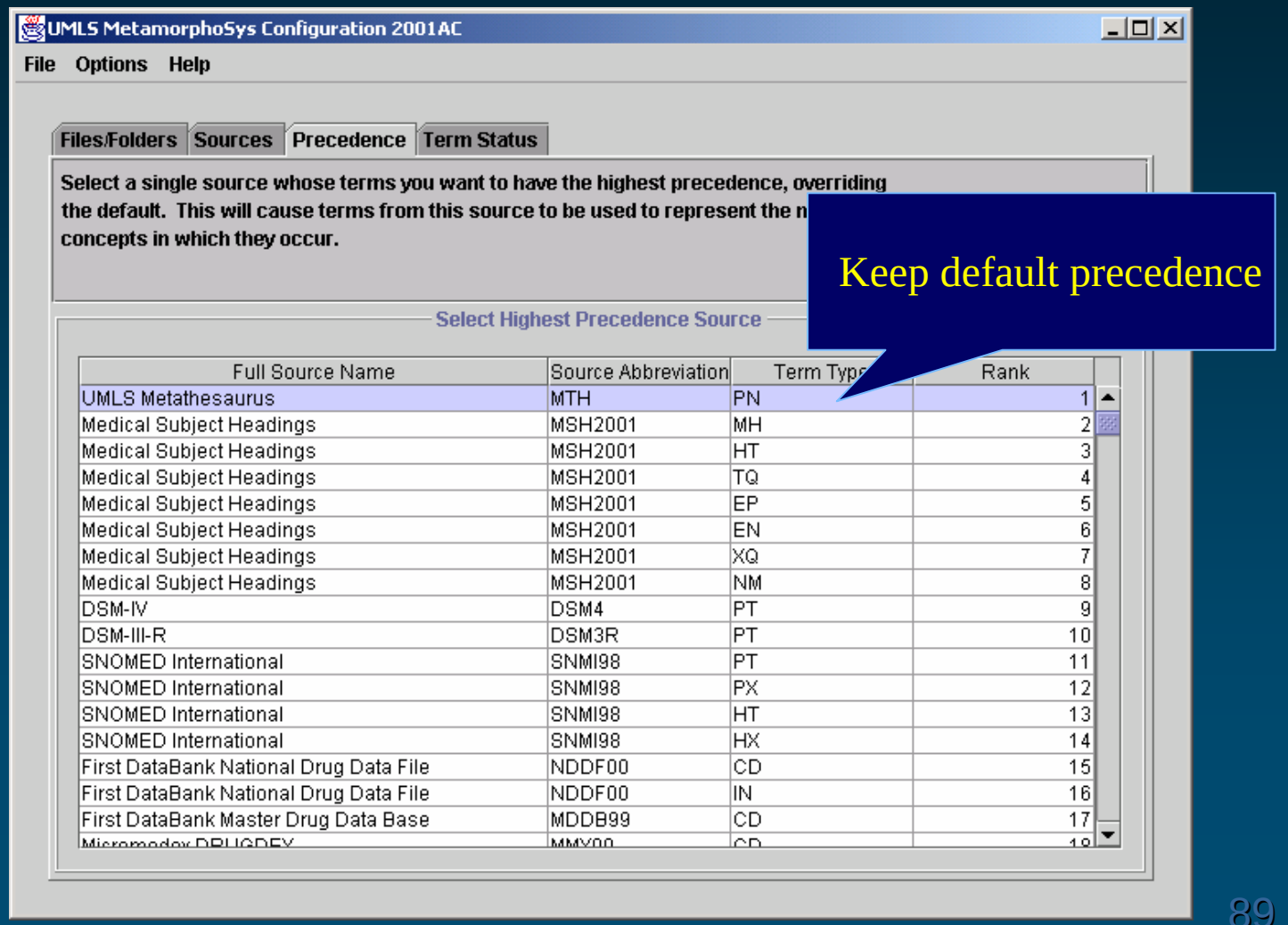
MRSO

```
C0001403|L0001403|S0010794|MSH2001|MH|D000224|0|
C0001403|L0001403|S0352253|CST95|GT|ADREN INSUFFIC|0|
C0001403|L0001403|S0352253|WHO97|IT|0410|2|
C0001403|L0001403|S0033587|MSH2001|PM|D000224|0|
C0001403|L0001403|S0469271|SNMI98|PT|DB-70620|3|
C0001403|L0367999|S0469267|SNMI98|SY|DB-70620|3|
C0001403|L0373744|S0471237|SNMI98|SY|DB-70620|3|
```

Adding to default suppressibility

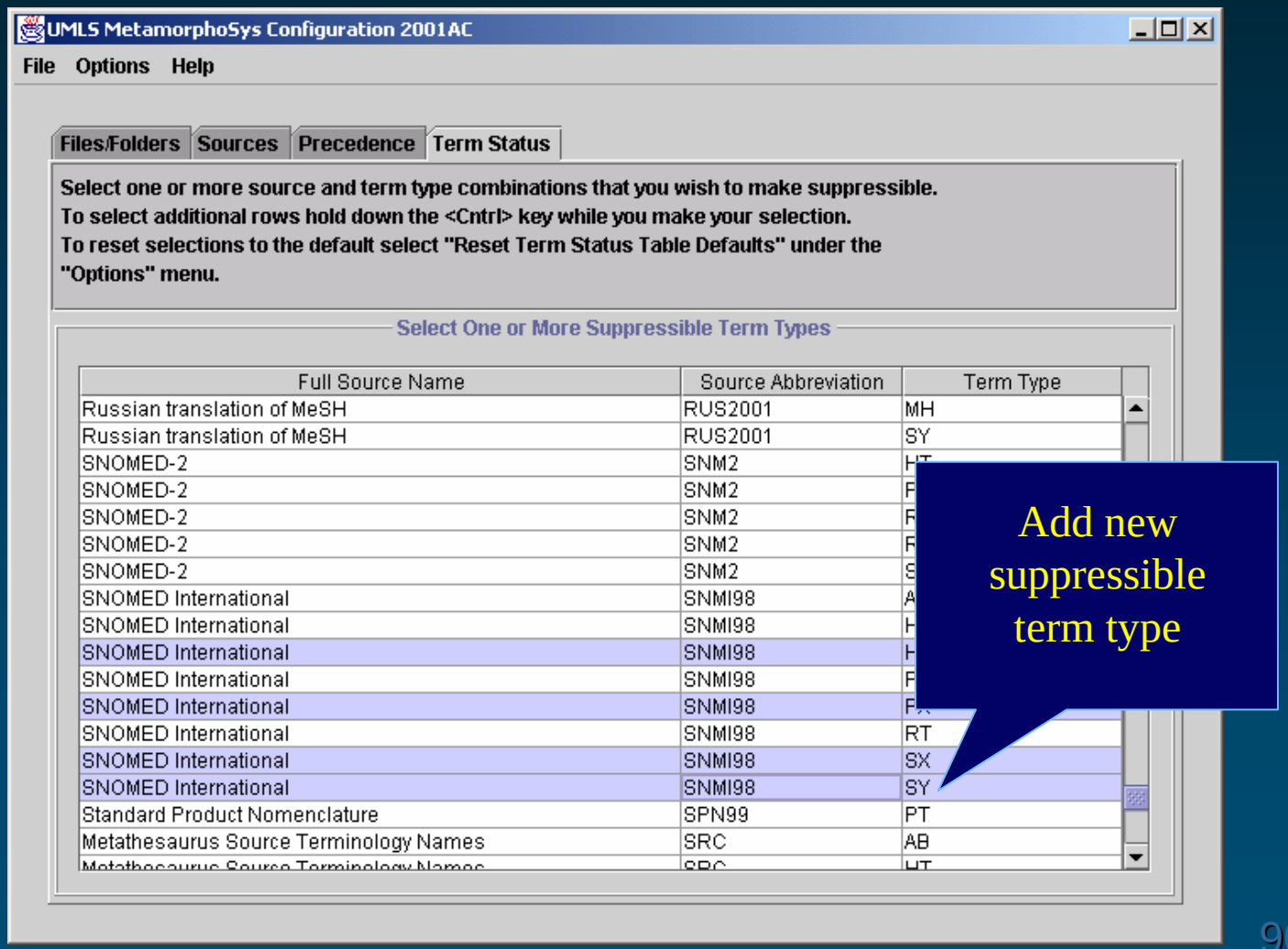


Adding to default suppressibility



NATIONAL LIBRARY OF MEDICINE

Adding to default suppressibility



Adding to default suppressibility

MRCON

```
C0001403|ENG|P|L0001403|PF|S0010794|Addison's Disease|0|
C0001403|ENG|P|L0001403|VC|S0352253|ADDISON'S DISEASE|0|
C0001403|ENG|P|L0001403|VO|S0033587|Disease, Addison|0|
C0001403|ENG|P|L0001403|VO|S0469271|Addison's disease, NOS|3|
C0001403|ENG|S|L0367999|PF|S0469267|Addison melanoderma|3|
C0001403|ENG|S|L0373744|PF|S0471237|Asthenia pigmentosa|3|
```

MRSO

```
C0001403|L0001403|S0010794|MSH2001|MH|D000224|0|
C0001403|L0001403|S0352253|CST95|GT|ADREN INSUFFIC|0|
C0001403|L0001403|S0352253|WHO97|IT|0410|2|
C0001403|L0001403|S0033587|MSH2001|PM|D000224|0|
C0001403|L0001403|S0469271|SNMI98|PT|DB-70620|3|
C0001403|L0367999|S0469267|SNMI98|SY|DB-70620|3|
C0001403|L0373744|S0471237|SNMI98|SY|DB-70620|3|
```

TS goes from “S” to “s”

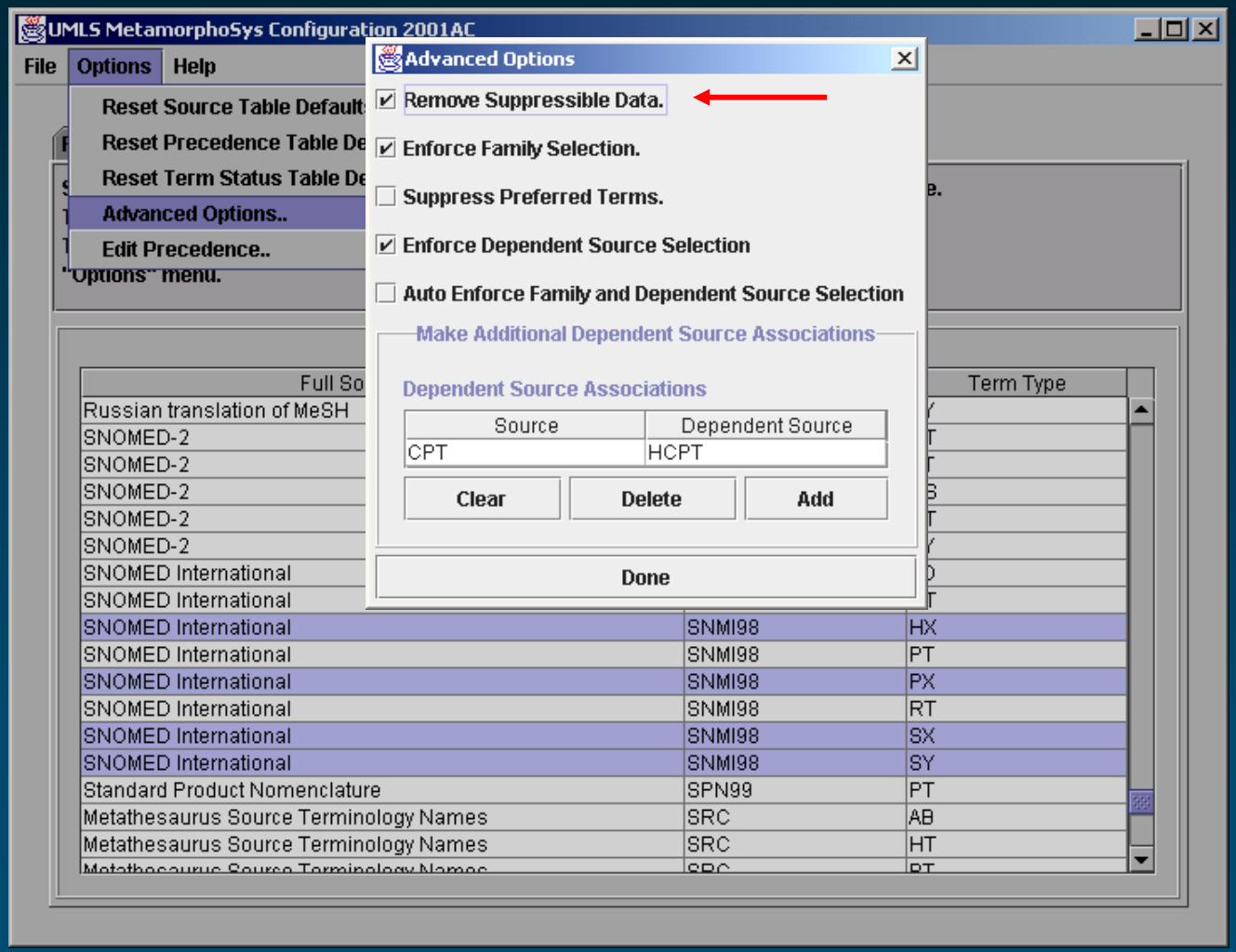
MRCON

```
C0001403|ENG|P|L0001403|PF|S0010794|Addison's Disease|0|
C0001403|ENG|P|L0001403|VC|S0352253|ADDISON'S DISEASE|0|
C0001403|ENG|P|L0001403|VO|S0033587|Disease, Addison|0|
C0001403|ENG|P|L0001403|VO|S0469271|Addison's disease, NOS|3|
C0001403|ENG|s|L0367999|PF|S0469267|Addison melanoderma|3|
C0001403|ENG|s|L0373744|PF|S0471237|Asthenia pigmentosa|3|
```

MRSO

```
C0001403|L0001403|S0010794|MSH2001|MH|D000224|0|
C0001403|L0001403|S0352253|CST95|GT|ADREN INSUFFIC|0|
C0001403|L0001403|S0352253|WHO97|IT|0410|2|
C0001403|L0001403|S0033587|MSH2001|PM|D000224|0|
C0001403|L0001403|S0469271|SNMI98|PT|DB-70620|3|
C0001403|L0367999|S0469267|SNMI98|SY|DB-70620|3|
C0001403|L0373744|S0471237|SNMI98|SY|DB-70620|3|
```

Removing suppressible data



Then, associated data are removed

MRCON

```
C0001403|ENG|P|L0001403|PF|S0010794|Addison's Disease|0|
C0001403|ENG|P|L0001403|VC|S0352253|ADDISON'S DISEASE|0|
C0001403|ENG|P|L0001403|VO|S0033587|Disease, Addison|0|
C0001403|ENG|P|L0001403|VO|S0469271|Addison's disease, NOS|3|
C0001403|ENG|s|L0367999|PF|S0469267|Addison melanoderma|3|
C0001403|ENG|s|L0373744|PF|S0471237|Asthenia pigmentosa|3|
```

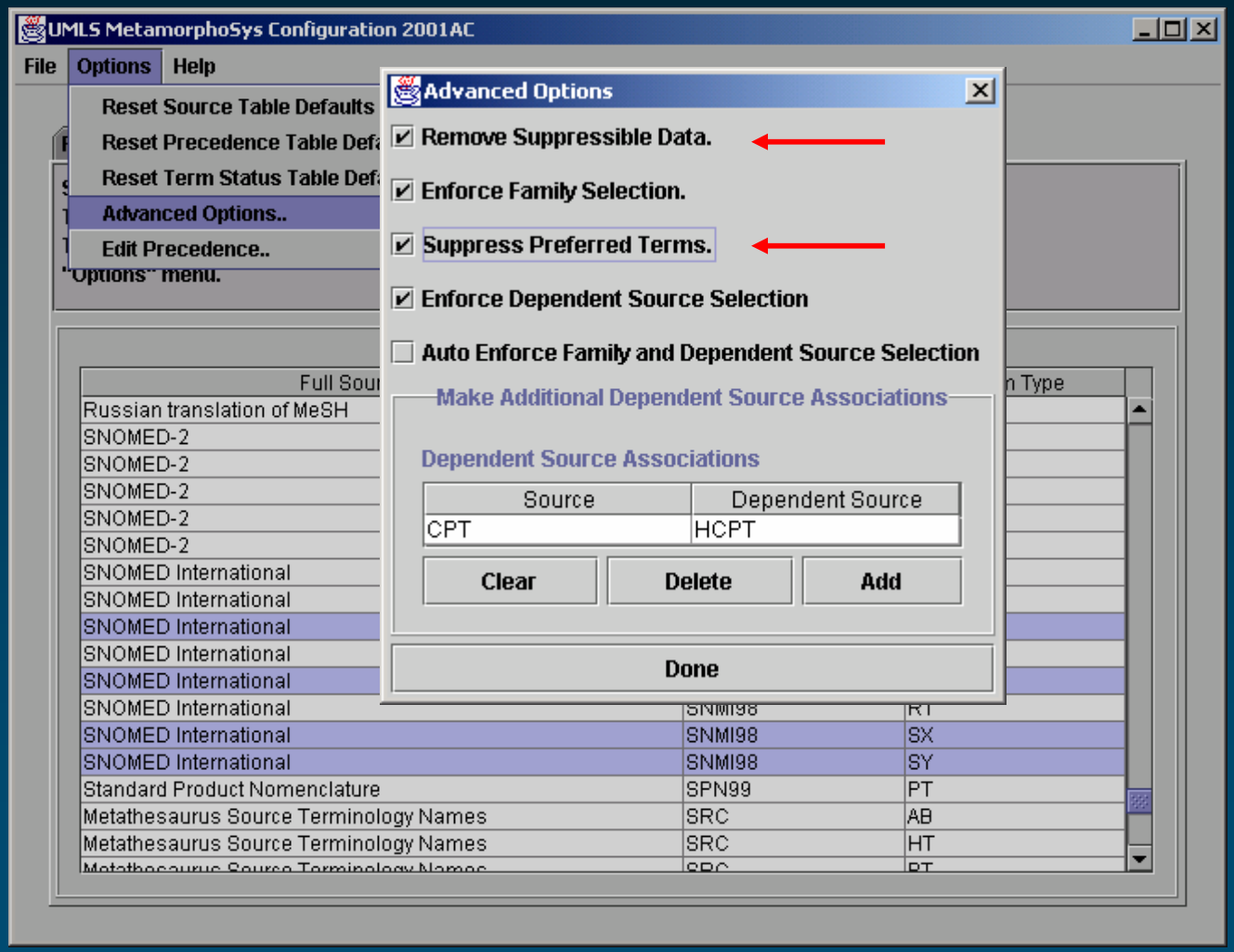
MRSO

```
C0001403|L0001403|S0010794|MSH2001|MH|D000224|0|
C0001403|L0001403|S0352253|CST95|GT|ADREN INSUFFIC|0|
C0001403|L0001403|S0352253|WHO97|IT|0410|2|
C0001403|L0001403|S0033587|MSH2001|PM|D000224|0|
C0001403|L0001403|S0469271|SNMI98|PT|DB-70620|3|
C0001403|L0367999|S0469267|SNMI98|SY|DB-70620|3|
C0001403|L0373744|S0471237|SNMI98|SY|DB-70620|3|
```

However, what if?

- Preferred name of concept comes from a suppressible source, term type?
- Concept needs a name, so the TS=P, STT=PF row is retained (there is no TS="p")

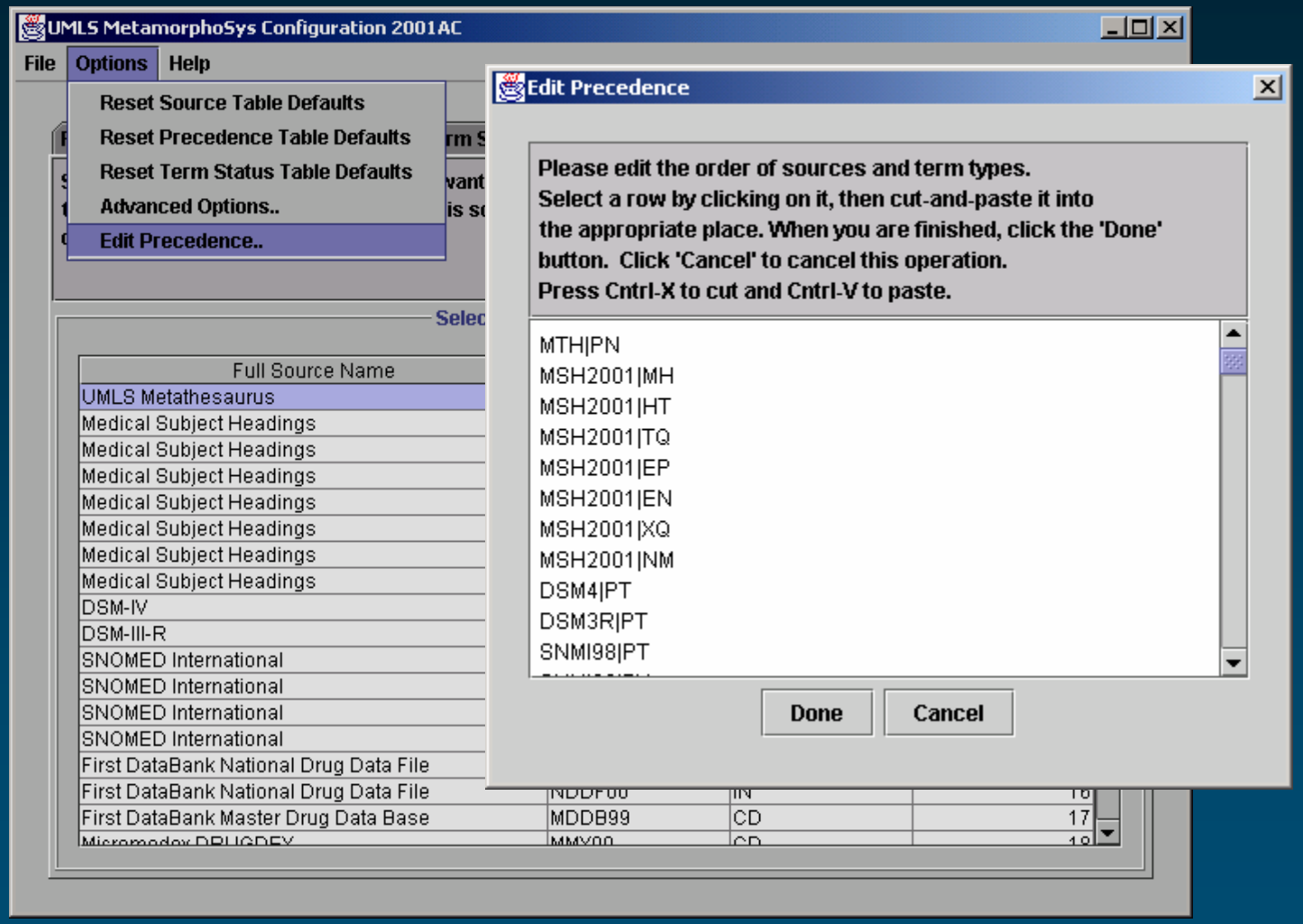
However, if both are selected..



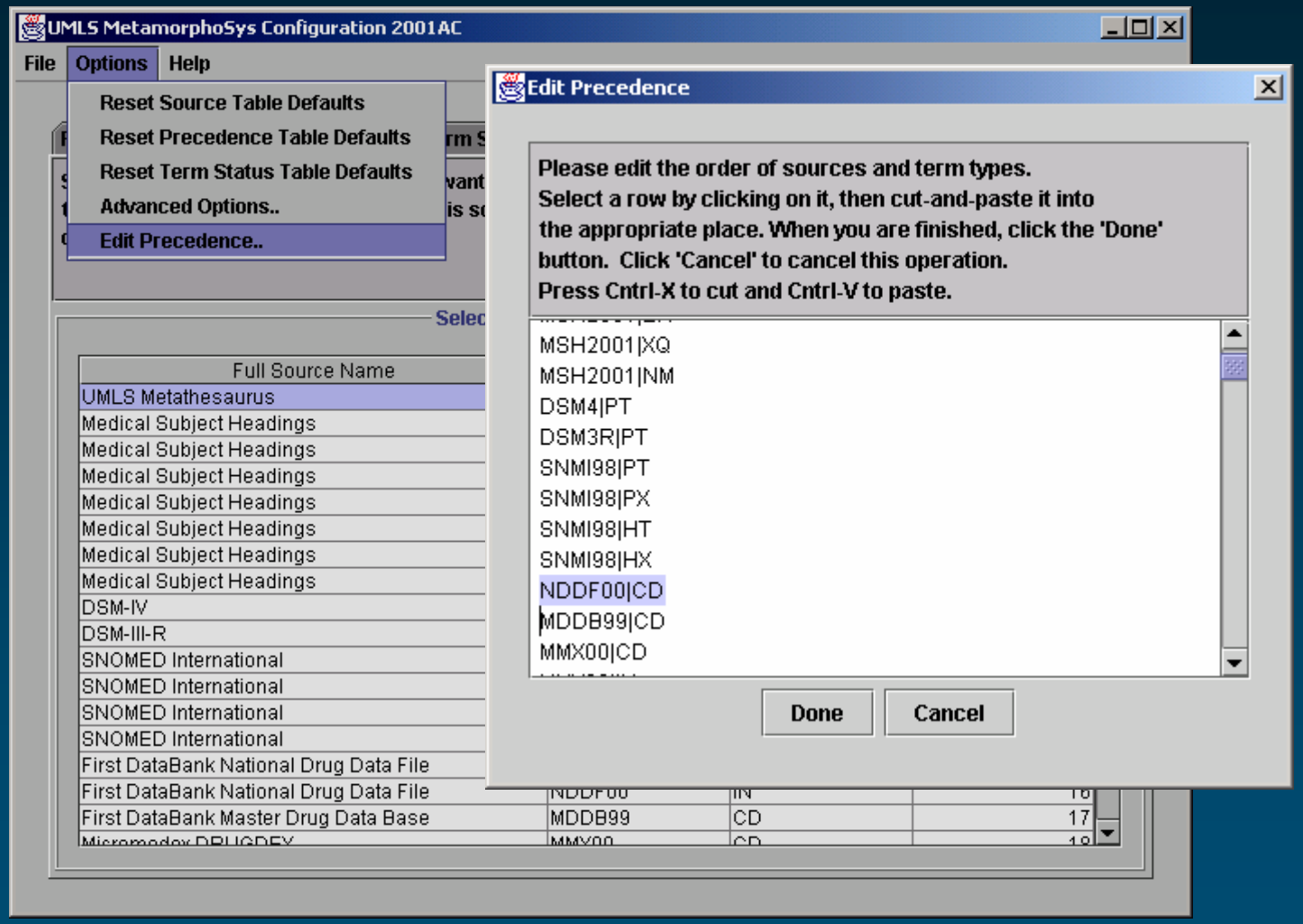
..and

- No other MRCON rows, or remaining MRCON rows are all suppressible,
- Then the entire concept is removed from the Metathesaurus (all files)

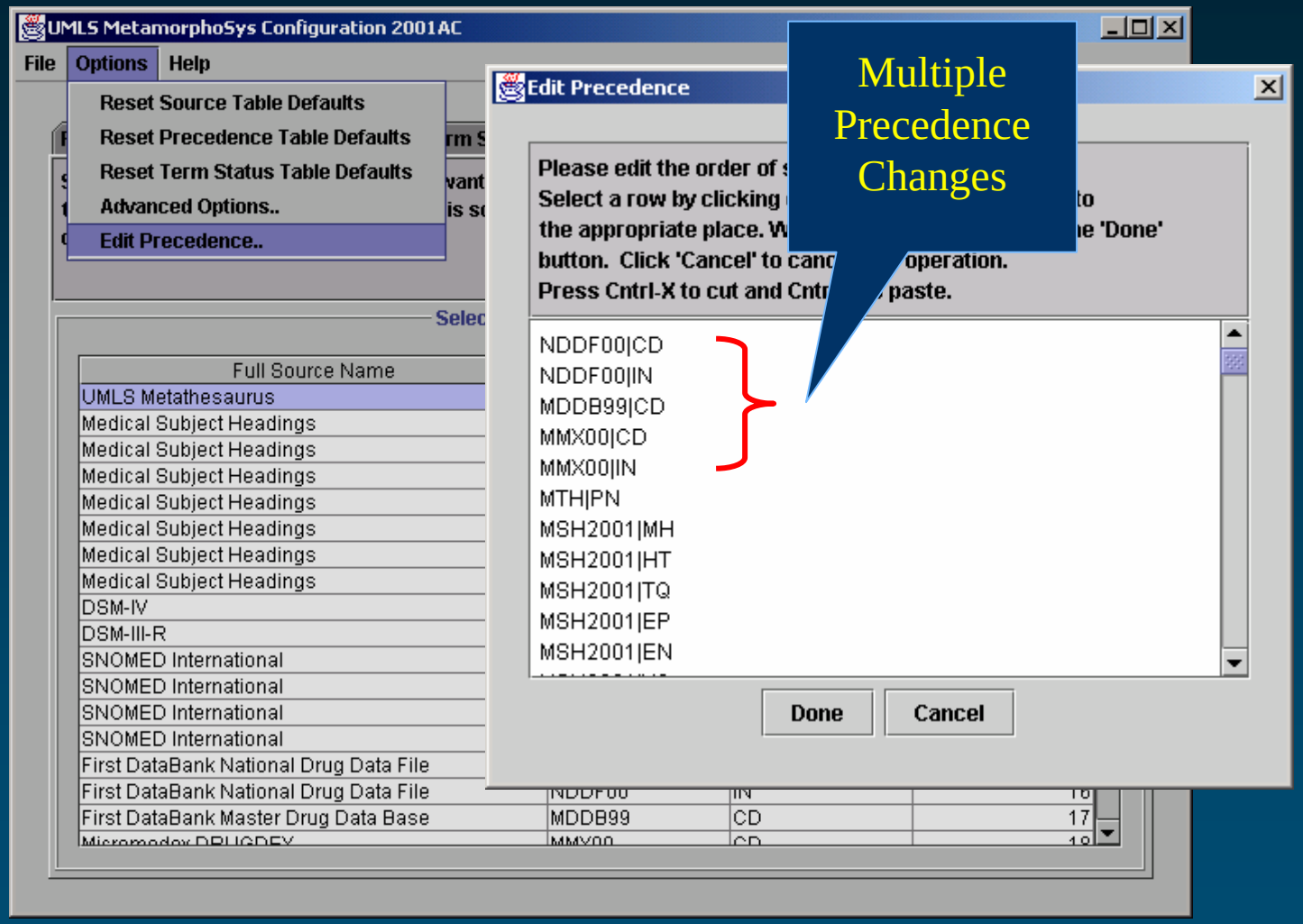
Editing precedence



Cut and Paste SAB/TTY



…to result in


UMLS MetamorphoSys Configuration 2001AC
File
Options
Help
Reset Source Table Defaults
Reset Precedence Table Defaults
Reset Term Status Table Defaults
Advanced Options..
Edit Precedence..
Edit Precedence
Please edit the order of
Select a row by clicking
the appropriate place.
button. Click 'Cancel' to
Press Cntrl-X to cut and
paste.
NDDF00|CD
NDDF00|IN
MDDB99|CD
MMX00|CD
MMX00|IN
MTH|PN
MSH2001|MH
MSH2001|HT
MSH2001|TQ
MSH2001|EP
MSH2001|EN
Done
Cancel
Multiple Precedence Changes
Full Source Name
UMLS Metathesaurus
Medical Subject Headings
DSM-IV
DSM-III-R
SNOMED International
First DataBank National Drug Data File
First DataBank Master Drug Data Base
MDDB99
CD
17

General comments on MetamorphoSys

- MetamorphoSys is configured to run with a specific release from its install directory – its use with other releases will cause unpredictable results
- MetamorphoSys propagates string-level suppressibility created and maintained by editors
- MetamorphoSys writes a log file (mmsys.log) in the subset directory that contains information about how that subset was generated
- STT computation better for some variants, still incomplete variants (e.g., VS) that need LVG

MetamorphoSys Configuration

- Program maintains the configuration as Java properties file
- **Do not** edit this file
- Can be saved for future runs
  - Default (*mmsys.prop.default*) should not be deleted
- Configuration is generic
  - Can be ported across versions of UMLS
  - Tied to source families, not just specific SABs
- All settings are saved (precedence, suppressibility)

Looking Ahead

- MetamorphoSys will become the “install” program for the UMLS Metathesaurus
- Customization by any axis: source, relationships, attributes
- Variety of output formats will be possible (Relational, XML, Atomic)
- MetamorphoSys will be able to act as an update client for the Metathesaurus

Outline of Tutorial

- Why customize? — Betsy Humphreys
- Metathesaurus basics — Olivier Bodenreider
- How to customize?
  - Customize sources (MetamorphoSys) — L. Roth & S. Srinivasan
  - Customize strings — Olivier Bodenreider
  - Customize synonyms
  - Customize relationships
  - Customize concept spaces
- Adding "local" terminology — Bill Hole

Beyond source-based customization

- More customization possible
  but
  - No tool available
  - Fits one specific purpose
  - Not necessarily useful for other purposes
  - No longer comparable with the original
  - New versions of the Metathesaurus need to be customized again

Using a model of the differences helps apply the customization systematically and effectively

NATIONAL LIBRARY OF MEDICINE

Beyond source-based customization

- Strings
- Synonyms
- Relationships: 3 different approaches
  - Semantic approach
  - Structural approach
  - Statistical approach
- Concept spaces

Overview of each section

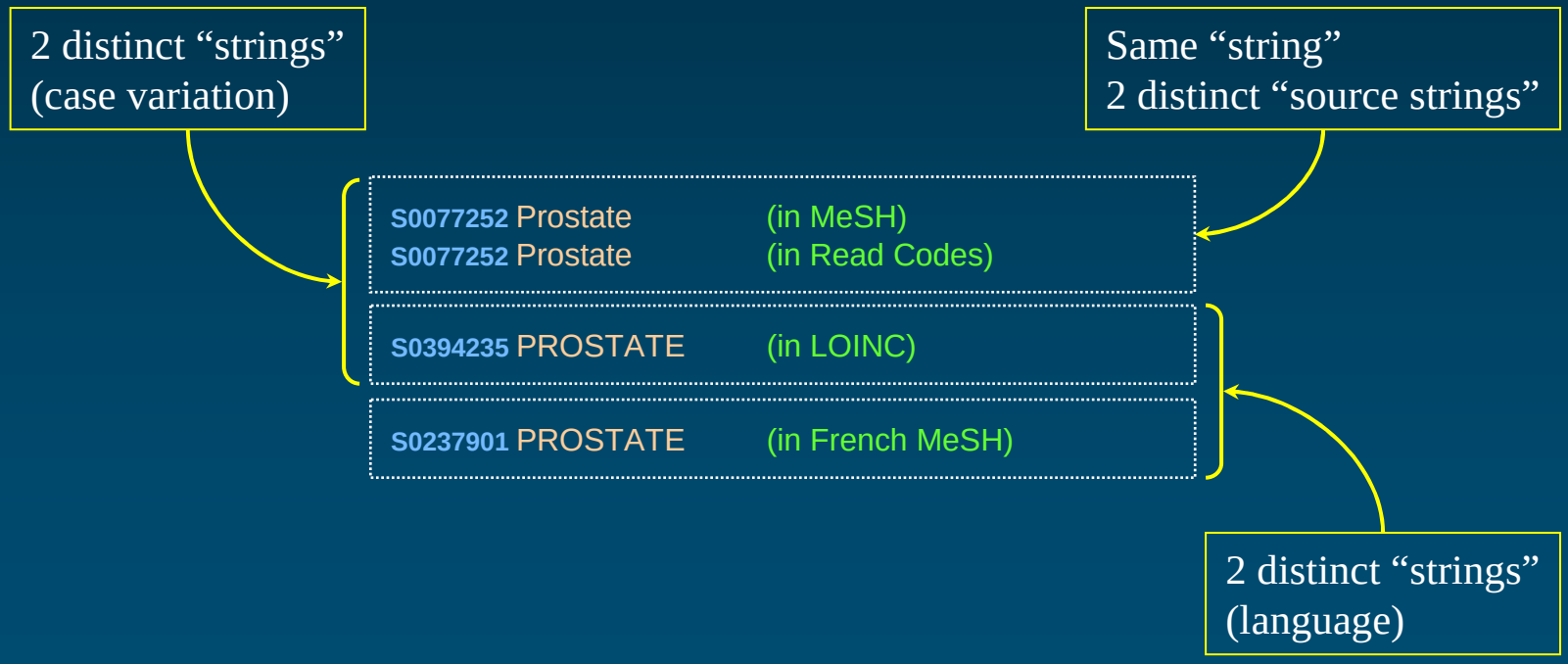
- Background
- Motivation
- Methods
- Example of use
- Discussion
  - Limitations
  - Alternative approaches

Outline of Tutorial

- Why customize? — Betsy Humphreys
- Metathesaurus basics — Olivier Bodenreider
- How to customize?
  - Customize sources (MetamorphoSys) — L. Roth & S. Srinivasan
  - Customize strings — Olivier Bodenreider
  - Customize synonyms
  - Customize relationships
  - Customize concept spaces
- Adding "local" terminology — Bill Hole

Background Strings

- Located in MRCON
- 1.9 million “source strings”



Background String attributes

- Language
- Preferred name in a source
- Lexical variants (case, inflection, word order, …)
- Other variants
  - Underspecification marker (Other, NOS)
  - Classification-specific marker (NEC)

Background More string attributes

- Source
- Term type (= type of string in a given source)
- Code in a given source
- Source-specific attributes
  - MN: Position in the hierarchy (MeSH)
  - SIC: ICD-9-CM code mapped to (SNOMED)
  - LFR: French name for a LOINC term
  - ICN: ICD-9-CM coding information
  - […]

Background Implicit string attributes

- Number of (families of) source vocabularies providing the string
- Presence in a target corpus

Motivation

- Reduce volume
- Select useful strings for natural language processing
- Select target-specific strings
- Filter out
  - Source-specific strings (e.g., truncated strings)
  - Purpose-specific strings (e.g., classification-specific strings, inverted terms)

Methods

- Identify string properties
- Combine the properties in order to create filters

Methods Identify string properties (1)

- Properties based on morphology (identified through regular expressions)
  - /, / for inverted terms 238,000
  - /[0-9]/ for strings containing digits 376,000
  - /^other|not elsewhere classified|NEC|without mention/ for classification feature 28,000
  - […]
  - Number of words in the string

Methods Identify string properties (2)

- Properties based on UMLS features

  - Redundancy: Number of (families of) source vocabularies providing this string 95,000
  - Term type (MRSO/TTY)
    - Chemical names 318,000
    - Branded drug names or supplies 62,000
    - Abbreviations and truncated strings 126,000

    […]
- Properties based on a corpus
  - e.g., strings found in MEDLINE 144,000

Methods Combine properties

- Using logical operators (AND, OR, NOT)
- 2 approaches
  - *A priori* model of the strings in a given context
  - Classification techniques against a target
- Traditional sensitivity/specificity balance

- e.g.: select English strings
  - Excluding chemical names
  - Excluding inverted terms
  - Found in more than one source vocabulary

Example of use

- Select UMLS strings useful for natural language processing

McCray A.T, Bodenreider O., Malley, J.D., Browne A.C.
Evaluating UMLS strings for natural language processing.
Proc AMIA Fall Symp. 2001 (in press) [S31 - Monday 2:00pm]

STR	NB_WORDS	ALLCAPS_ALWAYS	ALL_CLSP	ALL_UNSP	ANY_PARENTHETICAL	CT_COMMA_SPACE	CT_NON_ALPHANUM	CT_NUMBERS	CT_PUNCTUATION	CT_SYMBOLS	MI_AND_OR	NB_SOURCES	SUPPRESSIBLE_ALWAYS	TTY_CHEMICAL	TTY_LOINC	TTY_METADATA	TTY_PHRASE	TTY_PRESCRIPTION	TTY_SHORT_FORM
ADDISON DISEASE ✔	2											3							
Addison melanoderma	2											1							
Addisons Disease	2											2							
Addison's disease ✔	2											8							
Addison's disease NOS	3			x								1							
Addison's disease, NOS	3			x		x	x					1							
ADRENAL INSUFFICIENCY (ADDISON'S DISEASE)	4	x			x		x					1							
ADRENOCORTICAL INSUFFICIENCY, PRIMARY FAILURE	4	x				x	x					1							
Asthenia pigmentosa	2											1							
Bronzed disease	2											1							
DISEASE ADDISON'S	2	x										1							
Disease, Addison ✔	2					x	x					1							
Disease, Addisons	2					x	x					1							
Disease, Addison's ✔	2					x	x					1							
Disease;Addisons	2						x		x			1							
Melasma addisonii	2											1							
Primary adrenal deficiency	3											1							
Primary adrenocortical insuff	3											1	x						x
Primary adrenocortical insufficiency ✔	3											2							

Discussion

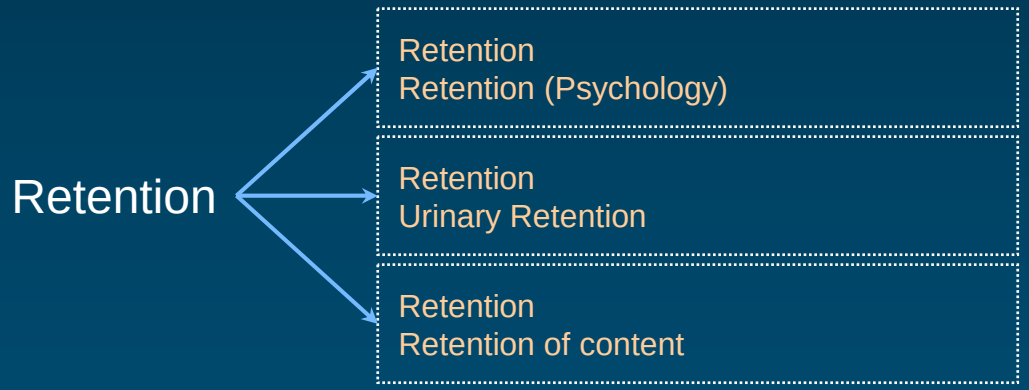
- Restricting to a given language is easier done through sources
- Filtering out strings may result in removing concepts
- Term status is relative to the preferred name, but does not identify the canonical form

Outline of Tutorial

- Why customize? — Betsy Humphreys
- Metathesaurus basics — Olivier Bodenreider
- How to customize?
  - Customize sources (MetamorphoSys) — L. Roth & S. Srinivasan
  - Customize strings — Olivier Bodenreider
  - Customize synonyms
  - Customize relationships
  - Customize concept spaces
- Adding “local” terminology — Bill Hole

Background

- Metathesaurus concepts are clusters of synonymous terms
- Polysemous terms may appear in more than one concept



Background

- Metathesaurus synonymy is not necessarily linguistic synonymy
  - Not fully specified terms
  - Granularity issues
  - Generic / prototypical

Prostate ✔ (in MeSH)
Prostatic gland

prostate ✘ (in COSTAR)
Prostatic Diseases

Prostate ✘ (in ICD-10)
Benign neoplasm of prostate

Posttransfusion hepatitis
Posttransfusion viral hepatitis

Asplenia
Congenital asplenia

Background

Myocardial Infarction

- Additionally, Metathesaurus synonyms include
  - Translated terms

    Infarctus du myocarde (French)
    Myocardinfarkt (German)
  - Lexical variants

    Myocardial Infarctions (plural)
    Infarction, Myocardial (permutation)
    Infarctions (Myocardial) (parentheses)
  - Acronyms

    MI
    MI - Myocardial infarction
  - Various kinds of terms (truncated, obsolete, …) as provided by source vocabularies

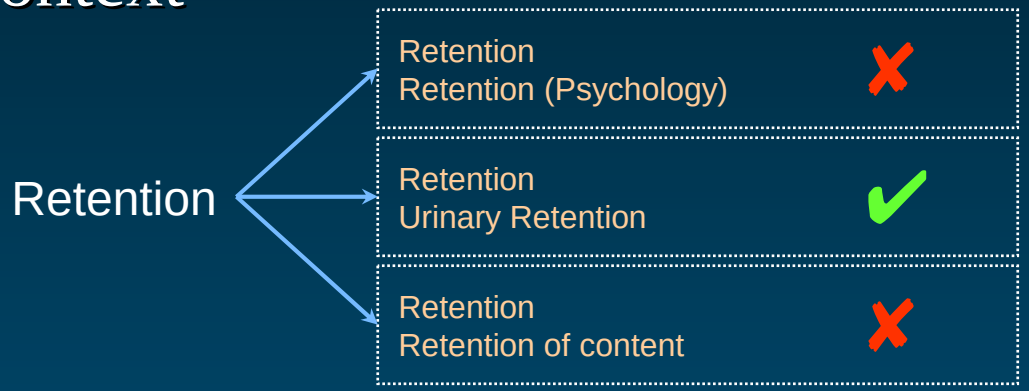
Background

- Some vocabularies implement their own notion of "synonymy"

depression and suicide	(preferred term)
suicide and depression	(synonym)
depression	(synonym)
suicide	(synonym)
cancer patients and suicide and depression	(synonym)
cancer patients and depression and suicide	(synonym)

Motivation

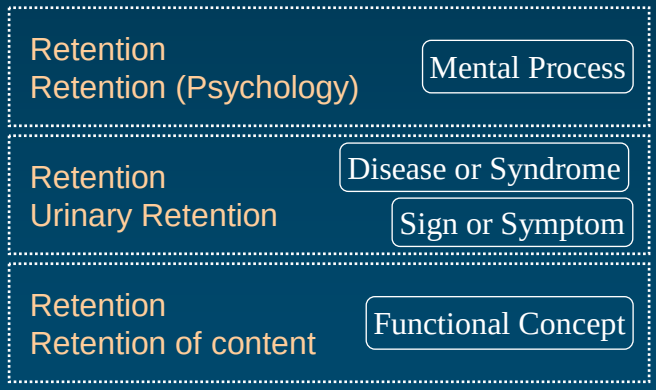
- Associate the right meaning with a string in a given context



- From the several strings associated with a meaning, select the most appropriate ones in a given context

Methods Associate the right meaning

- Use the “suppressible synonym” flag
  - Identifies not fully specified names
  - A fully specified name usually exists among the synonyms (sometimes created by NLM)
- Restrict the domain
  - In order to limit polysemy
  - Implies
    - A priori knowledge
    - Interaction with users
- Word sense disambiguation research area



Methods Most appropriate strings

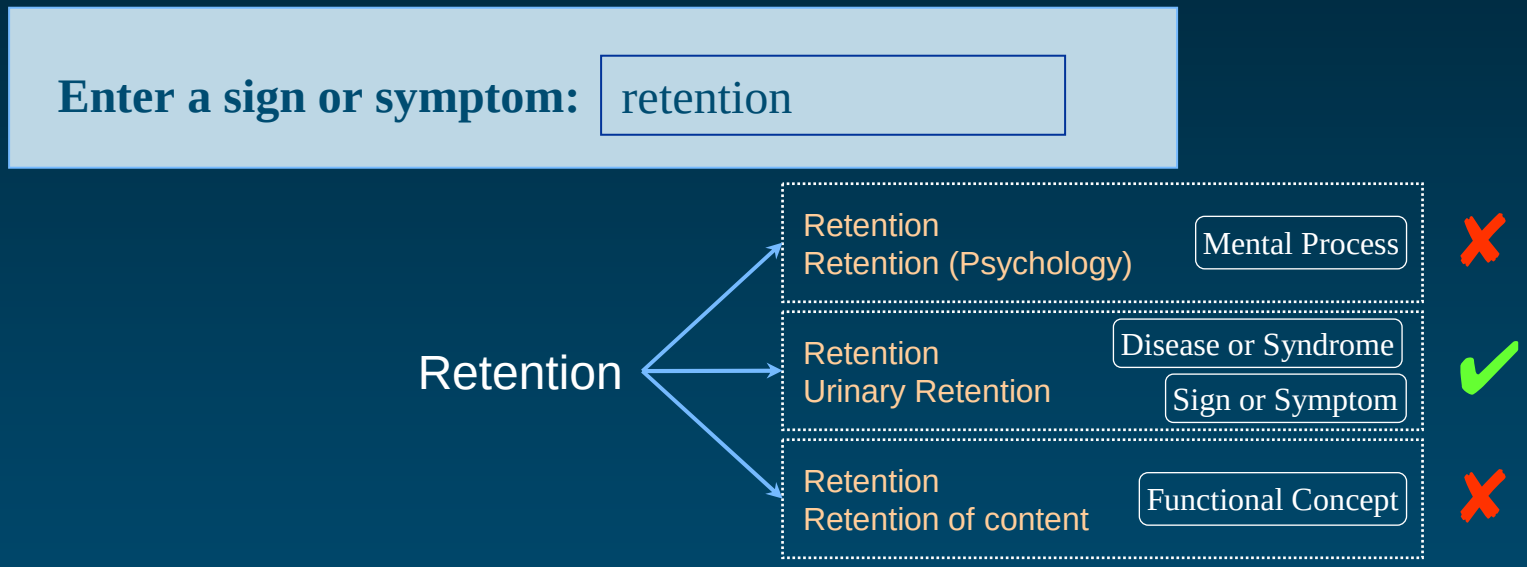
- Recognize and filter out lexical variants
  - Canonical form
  - Normalization
- Filter against a corpus
  - To find the most common form in your target

MEDLINE 1999

Fallen arch	
Fallen arches	
Flat foot NOS	
Flat Feet	✔
Flatfeet	✔
Flatfoot	✔
Foot, Flat	
Low medial arch of foot	
Pes Planus	✔
Pes planovalgus	✔
Pes valgus	✔

Example of use

- Disambiguate according to the context



- Filter redundant lexical variants from a list of terms in a Metathesaurus concept

Discussion

- Word sense disambiguation
  - Never trivial
  - Still open research area (linguistics)
  - Often involves statistical analysis of the context
- The Metathesaurus partially addresses the issue of not fully specified terms

Outline of Tutorial

- Why customize? — Betsy Humphreys
- Metathesaurus basics — Olivier Bodenreider
- How to customize?
  - Customize sources (MetamorphoSys) — L. Roth & S. Srinivasan
  - Customize strings — Olivier Bodenreider
  - Customize synonyms
  - Customize relationships
  - Customize concept spaces
- Adding “local” terminology — Bill Hole

Customize relationships

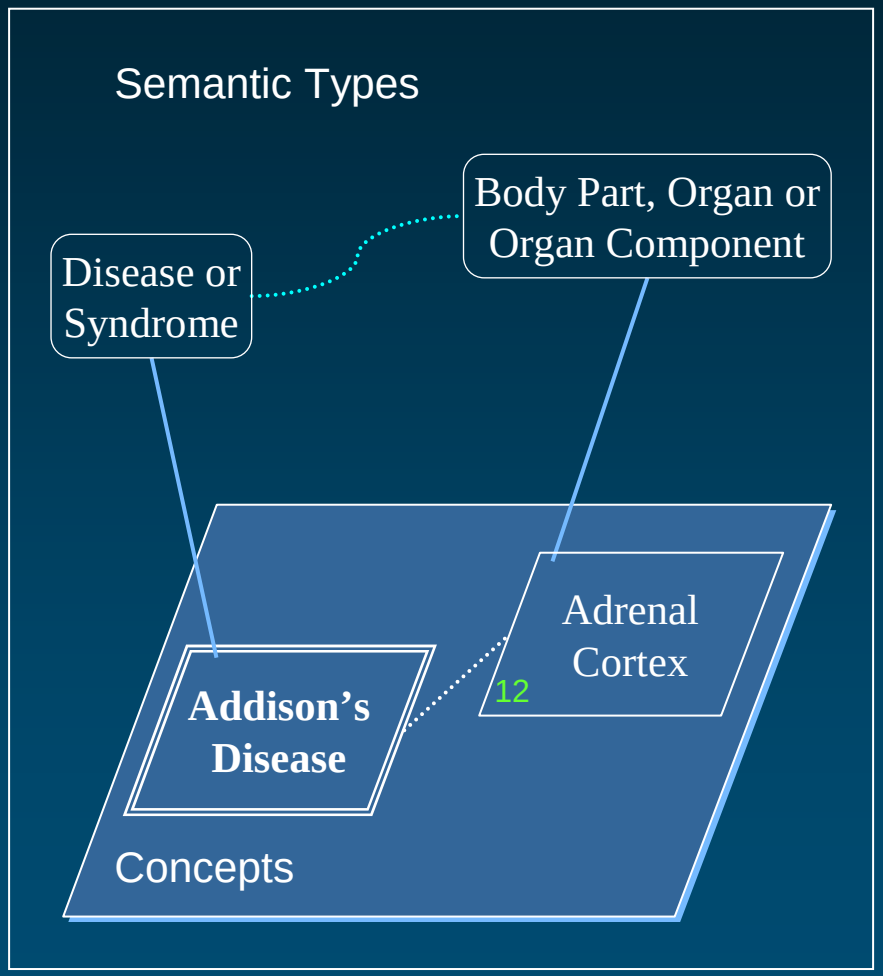
- With reference to the Semantic Network (semantic approach)
- Hierarchical relationships (structural approach)
- Co-occurrences (statistical approach)

Customize Relationships

1. Semantic Approach

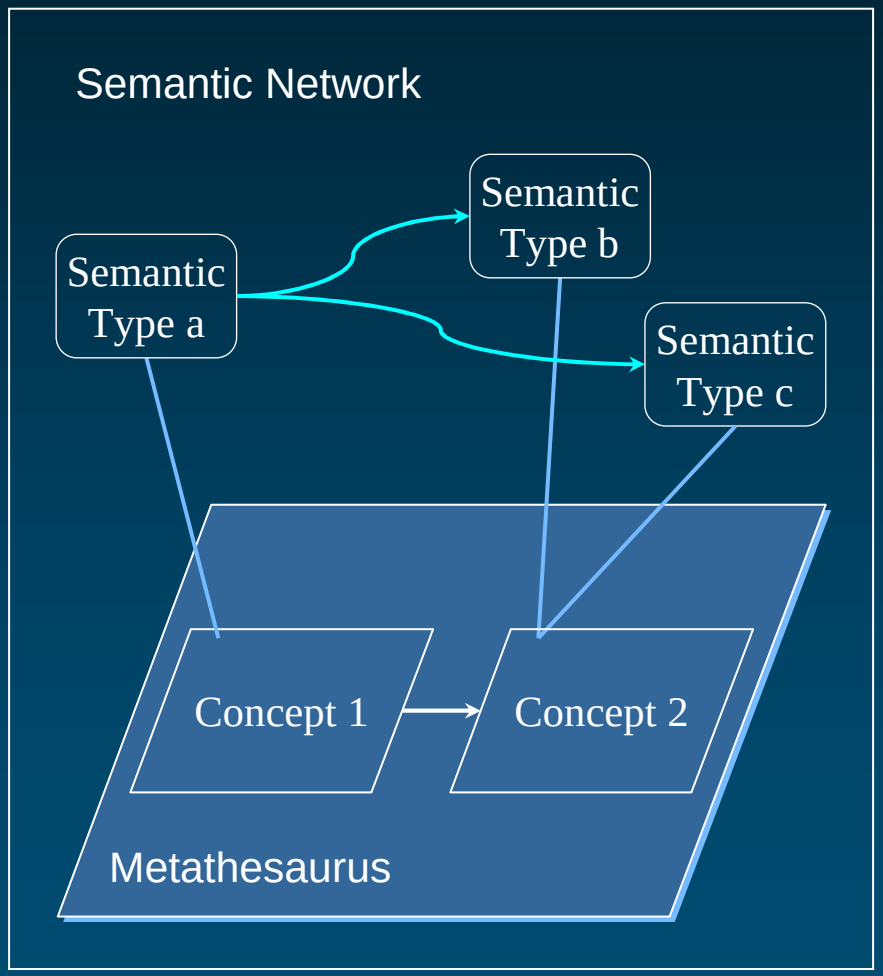
Background UMLS structure (nodes)

- Two-level structure
  - Semantic Network (134 semantic types)
  - Metathesaurus (800,000 concepts)



Background UMLS structure (links)

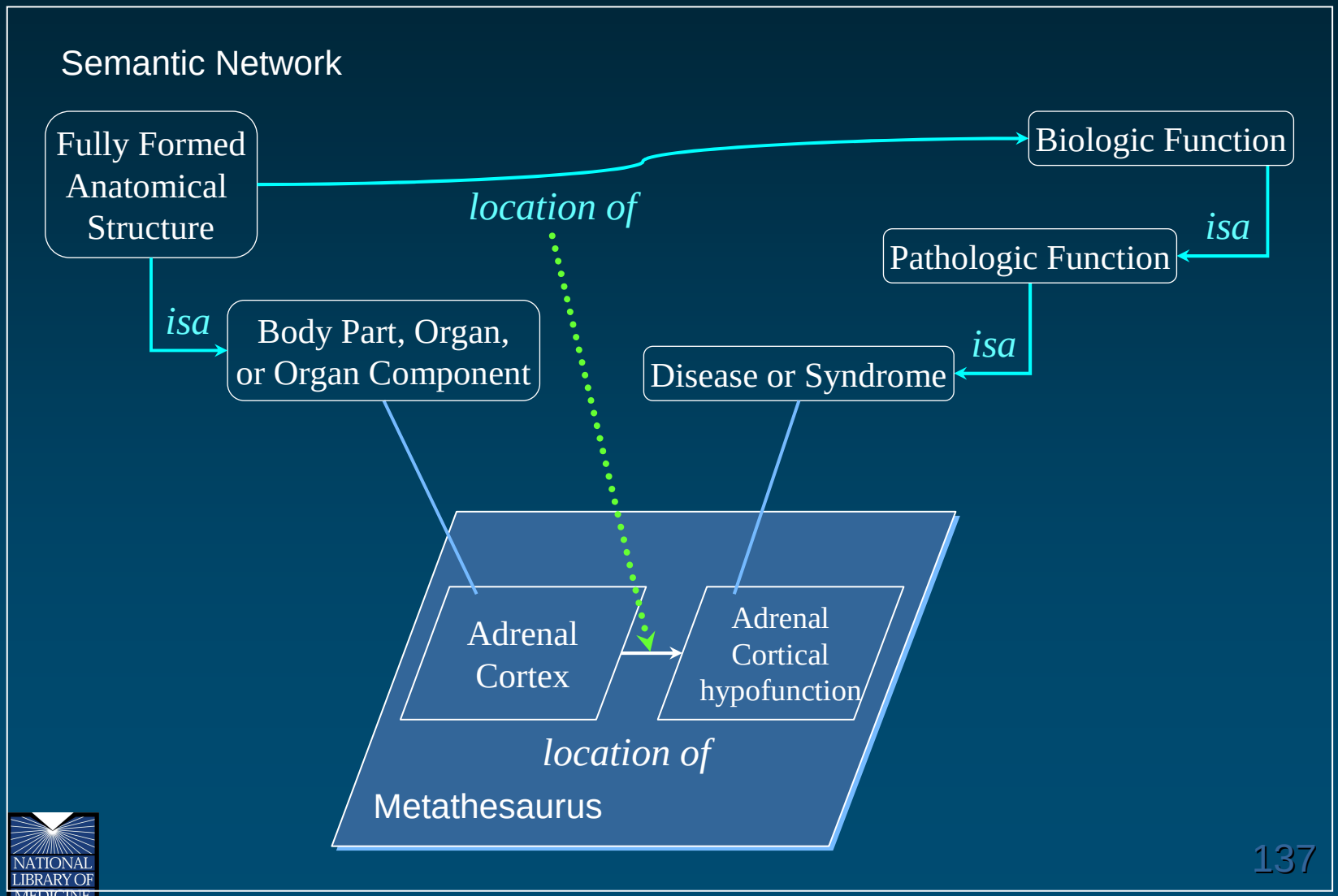
- Semantic network relationships
- Categorization
- Interconcept relationships



Background UMLS structure (links)

- Semantic network relationships
  - Hierarchical or associative
  - General (definitional) knowledge
  - May or may not hold at the concept level
- Categorization
  - Links each concept to (at least) one broad category
  - Either *isa* or *is an instance of* relationships
- Interconcept relationships
  - Hierarchical, associative or statistical
  - Factual knowledge

Relationships can inherit semantics



Motivation

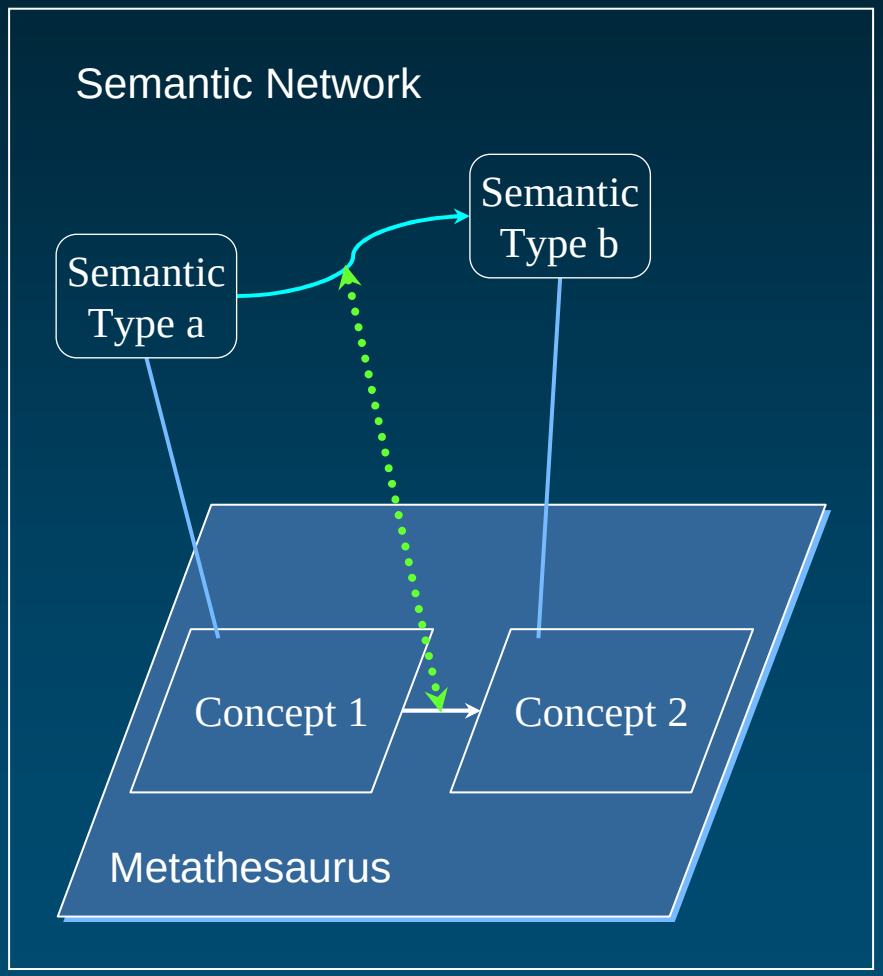
- Check the consistency of the two levels
  - Semantic network
  - Metathesaurus
- Check the consistency between
  - Semantic network relationships
  - Interconcept relationships
- Discrepancies may indicate
  - Inaccurate relationship
  - Inaccurate categorization

Motivation

- More generally
  - The Semantic Network represents some kind of upper-level ontology of the biomedical domain
  - The organization of Metathesaurus concepts
    - is *expected* to be compatible with the upper level
    - is *required* to be compatible with the upper level if reasoning is to be supported

Methods

- For each pair of related concepts
  - Get their semantic types
  - Get all the "expanded" semantic network relationships between the two semantic types (transitive closure)
  - Compare
    - Interconcept relationship
    - Sem. Net. relationships



Methods

- Possible outcome
  - ICR = SNR → validate
  - ICR descendant of SNR → validate
  - ICR and SNR not compatible → reject
  - Unspecified ICR (no RELA) → infer/reject
  - ICR not in the Semantic Network

ICR: Inter-concept relationship
SNR: Semantic Network relationship

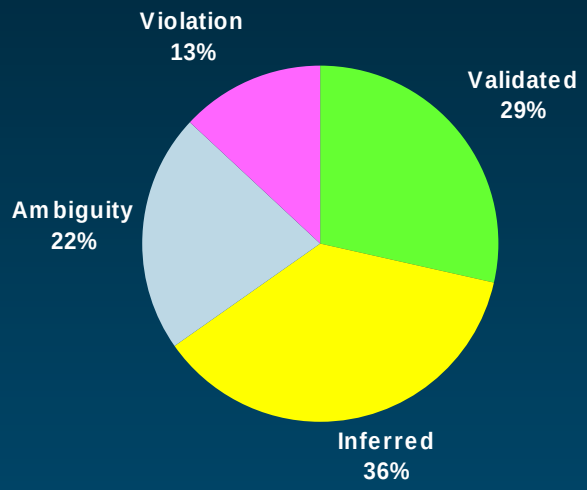
Example of use

- Validate, infer or reject interconcept relationships by comparison to the relationships defined between the semantic types assigned to the concepts

McCray A.T, Bodenreider O.
A conceptual framework for the biomedical domain.
in Sung, M. and Green, R. eds. *Semantics of Relationships*, Kluwer, 2001, (in press).

Example of use Results

- 6894 interconcept relationships
  - among the 3764 concepts in the semantic neighborhood of “Heart”



Discussion

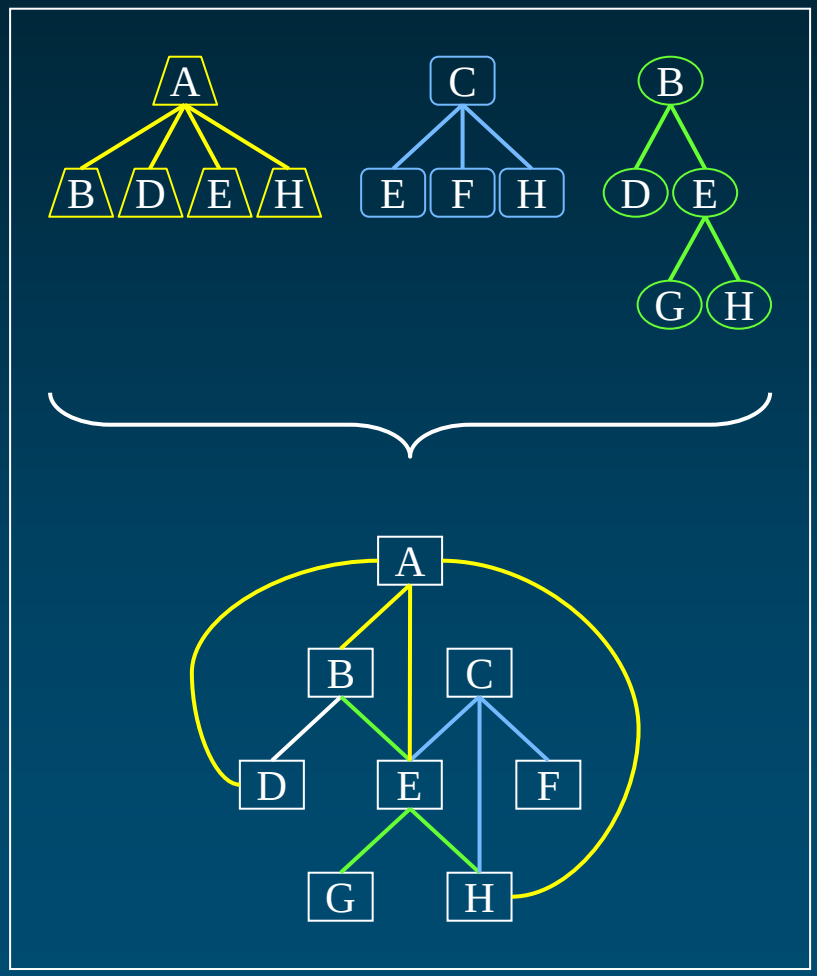
- Interconcept relationships recorded in the Metathesaurus are not censored
- The Semantic Network
  - Provides semantic constraints
  - Can be used to select Metathesaurus relationships that are “semantically sound”
- Limitations
  - Ambiguous SN relationships
  - Unspecified Metathesaurus relationships
  - Need for some manual review

Customize Relationships

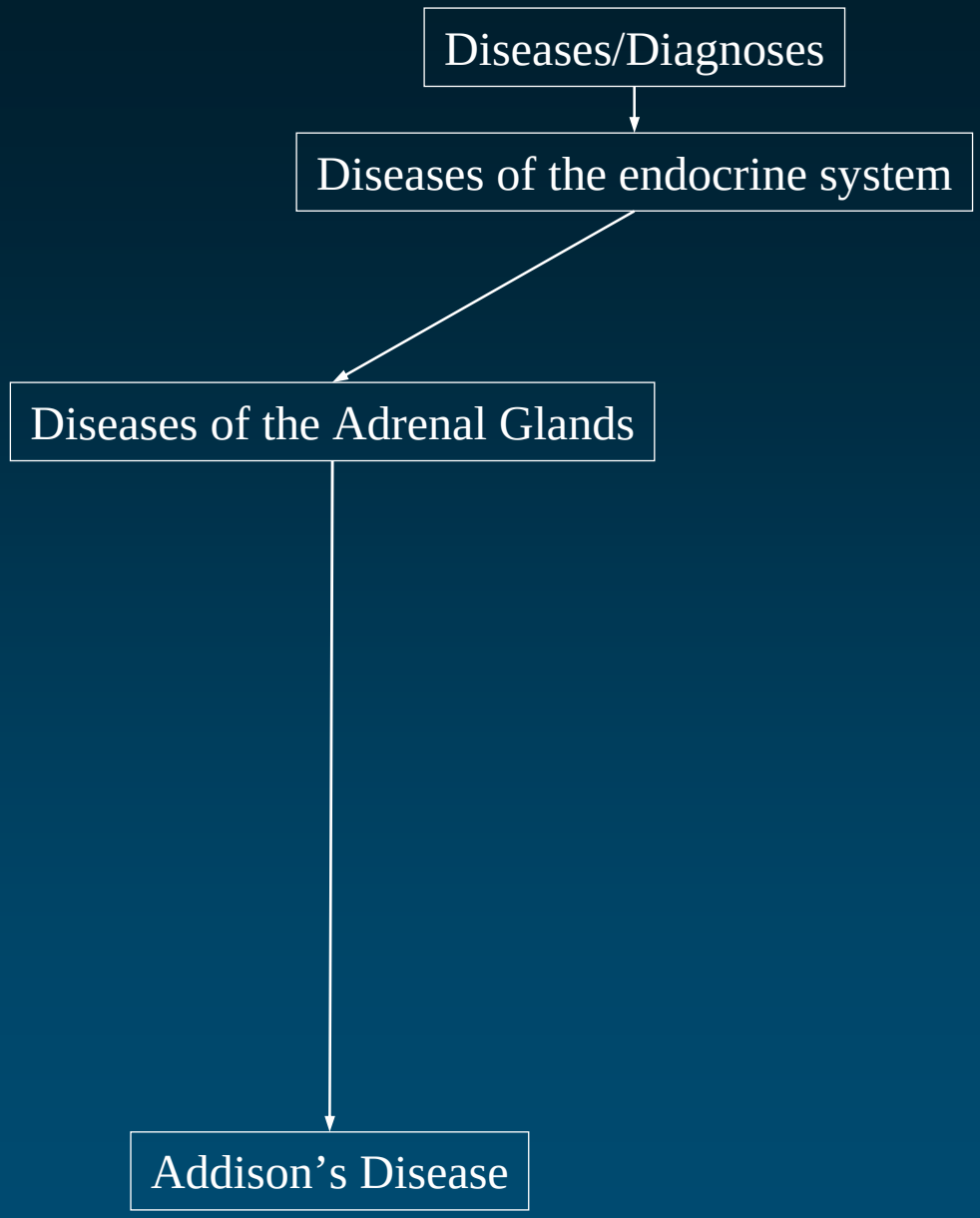
❷ Structural Approach

Background

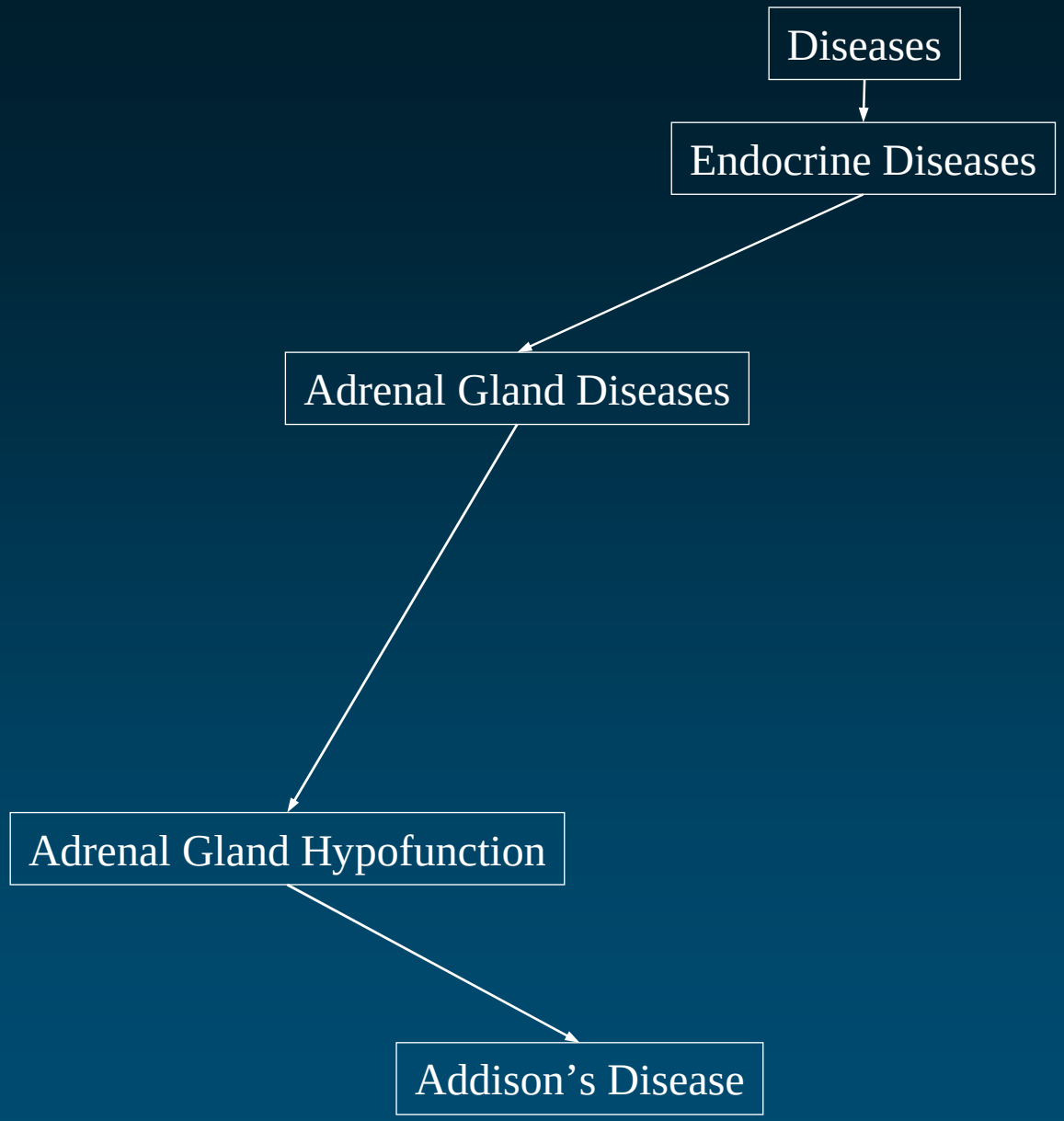
- The Metathesaurus is often seen as a bunch of trees
- Trees can be combined into a (directed) graph
- Hierarchies (esp. taxonomies) are based on partial ordering relationship
- Hierarchical relationships in the Metathesaurus are expected to result in a Directed Acyclic Graph (DAG)



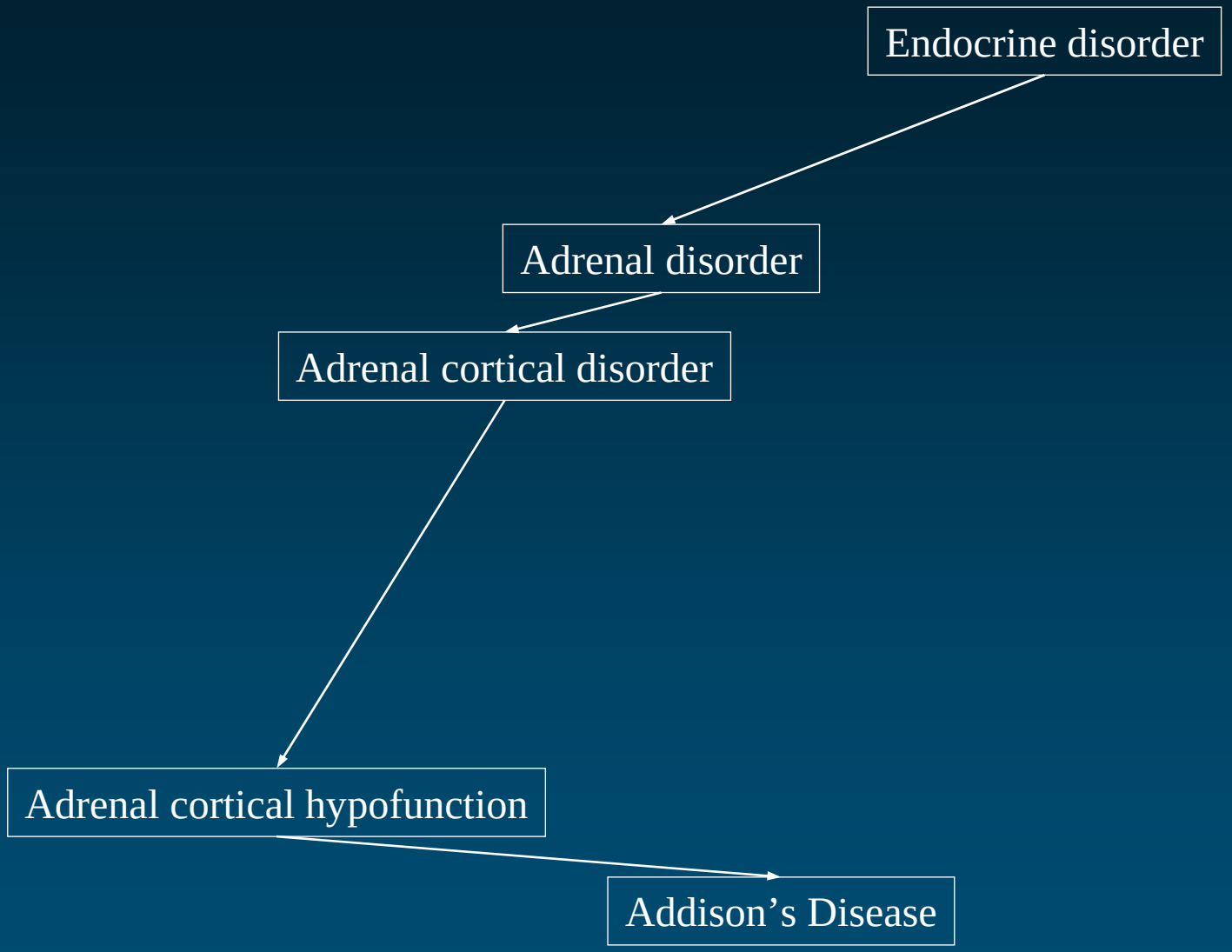
SNOMED International *tree*



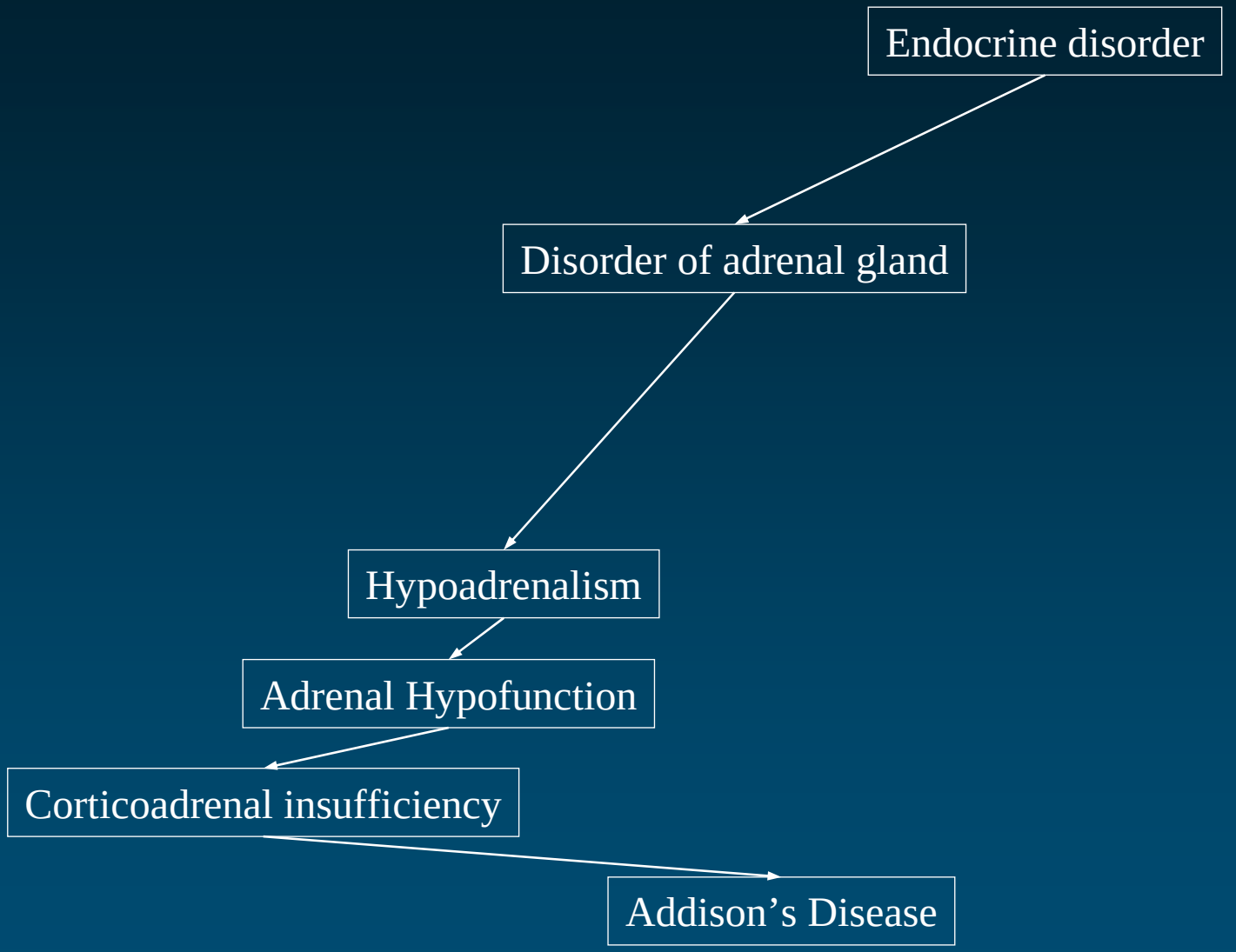
MeSH *tree*


Diseases
Endocrine Diseases
Adrenal Gland Diseases
Adrenal Gland Hypofunction
Addison’s Disease

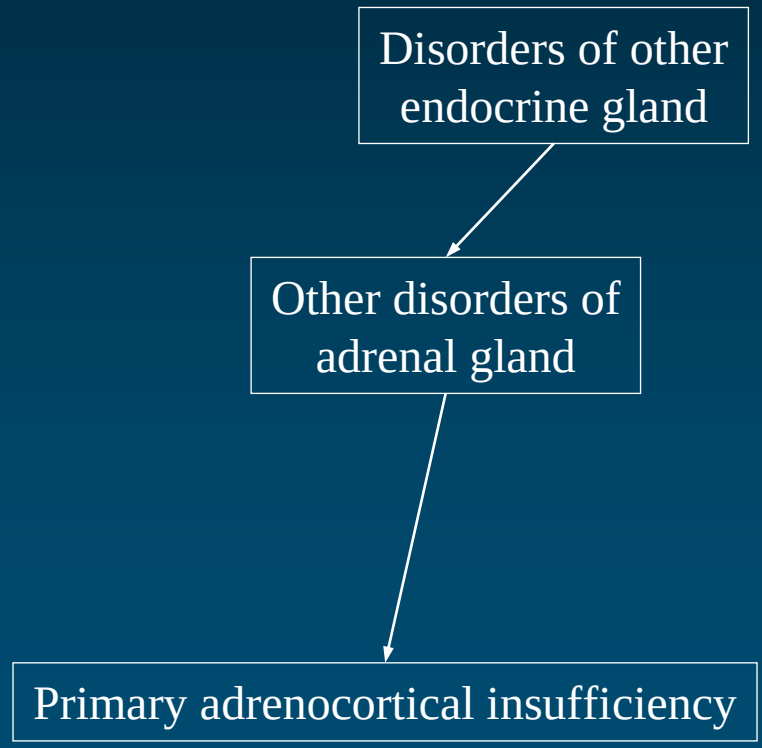
AOD *tree*


Endocrine disorder
Adrenal disorder
Adrenal cortical disorder
Adrenal cortical hypofunction
Addison’s Disease

Read Codes *tree*



ICD-10 *tree*


Disorders of other endocrine gland
Other disorders of adrenal gland
Primary adrenocortical insufficiency

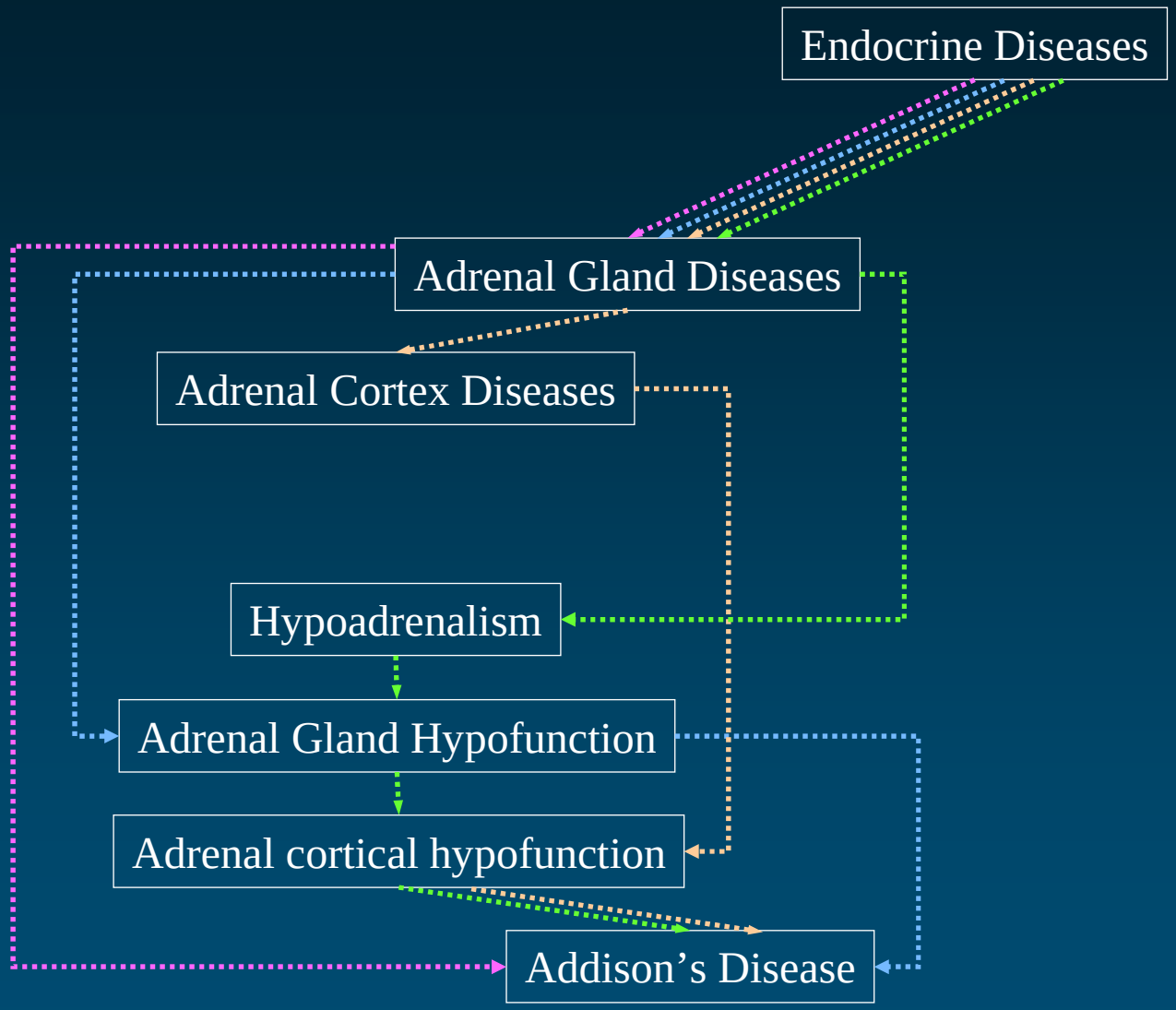
Metathesaurus *graph*

SNOMED

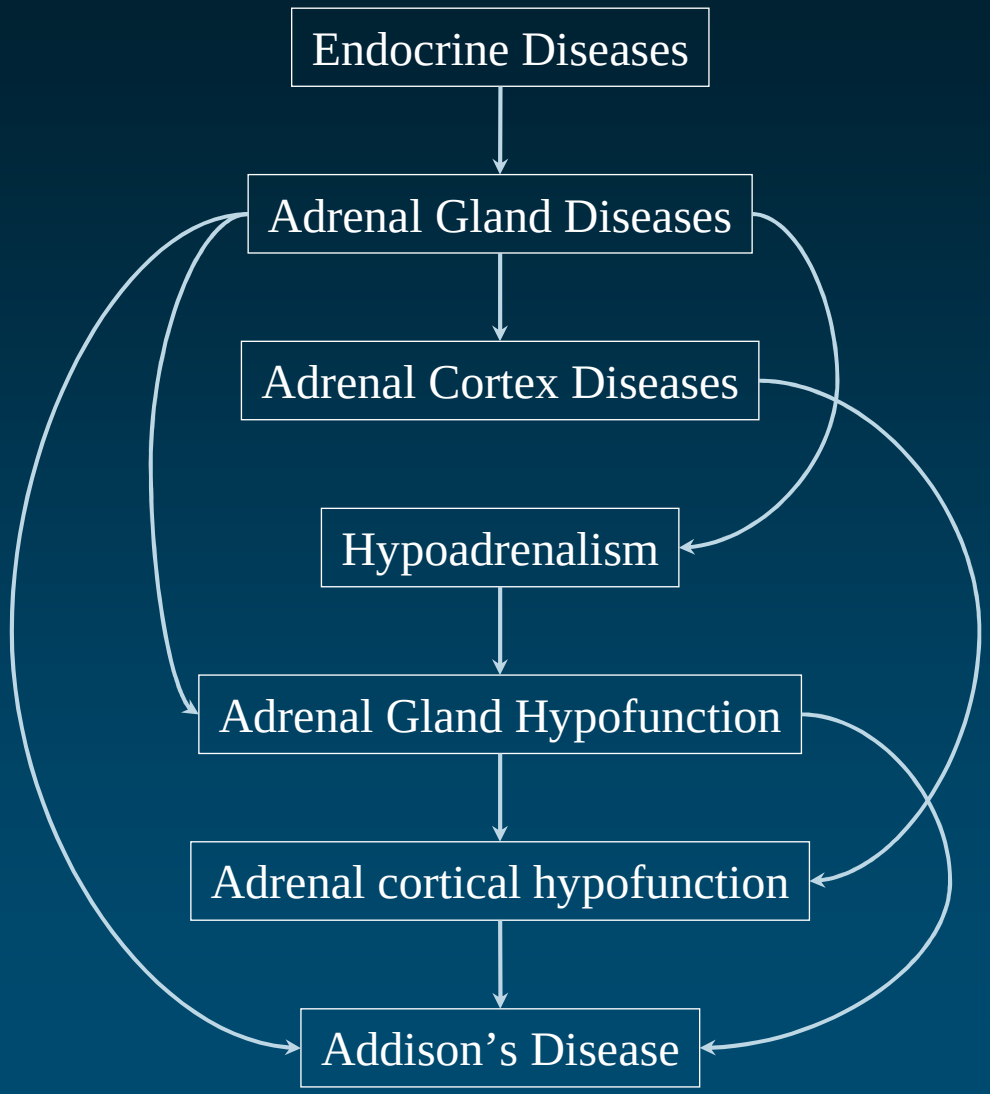
MeSH

AOD

Read Codes

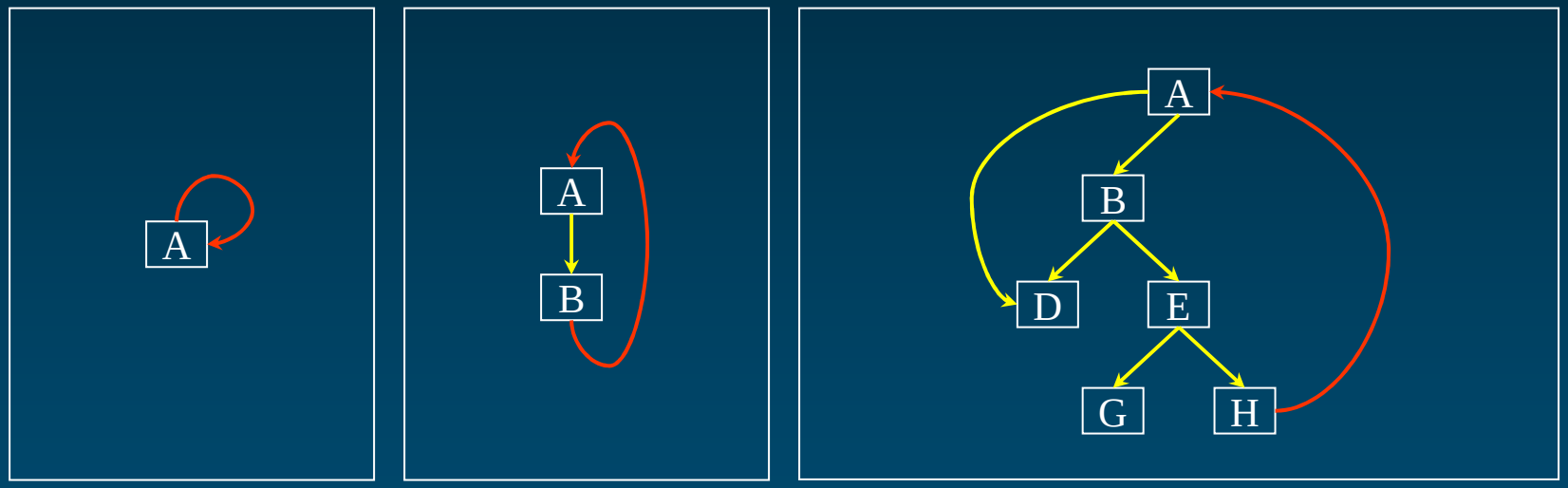


Metathesaurus *graph*



Circular hierarchical relationships

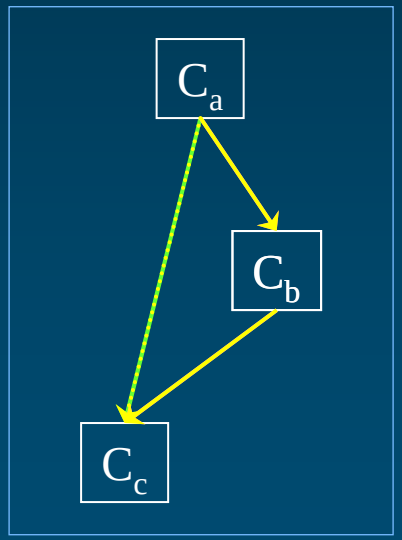
"back edge" from a child concept to a parent concept



Reflexive Direct Indirect

Motivation

- Circular hierarchical relationships are indicative of potential semantic issues
  - Wrong relationships
  - Non-hierarchical “hierarchical” relationships
- Some graph operations cannot be performed unless graph is acyclic
  - Transitive reduction



Methods

- Identify cycles

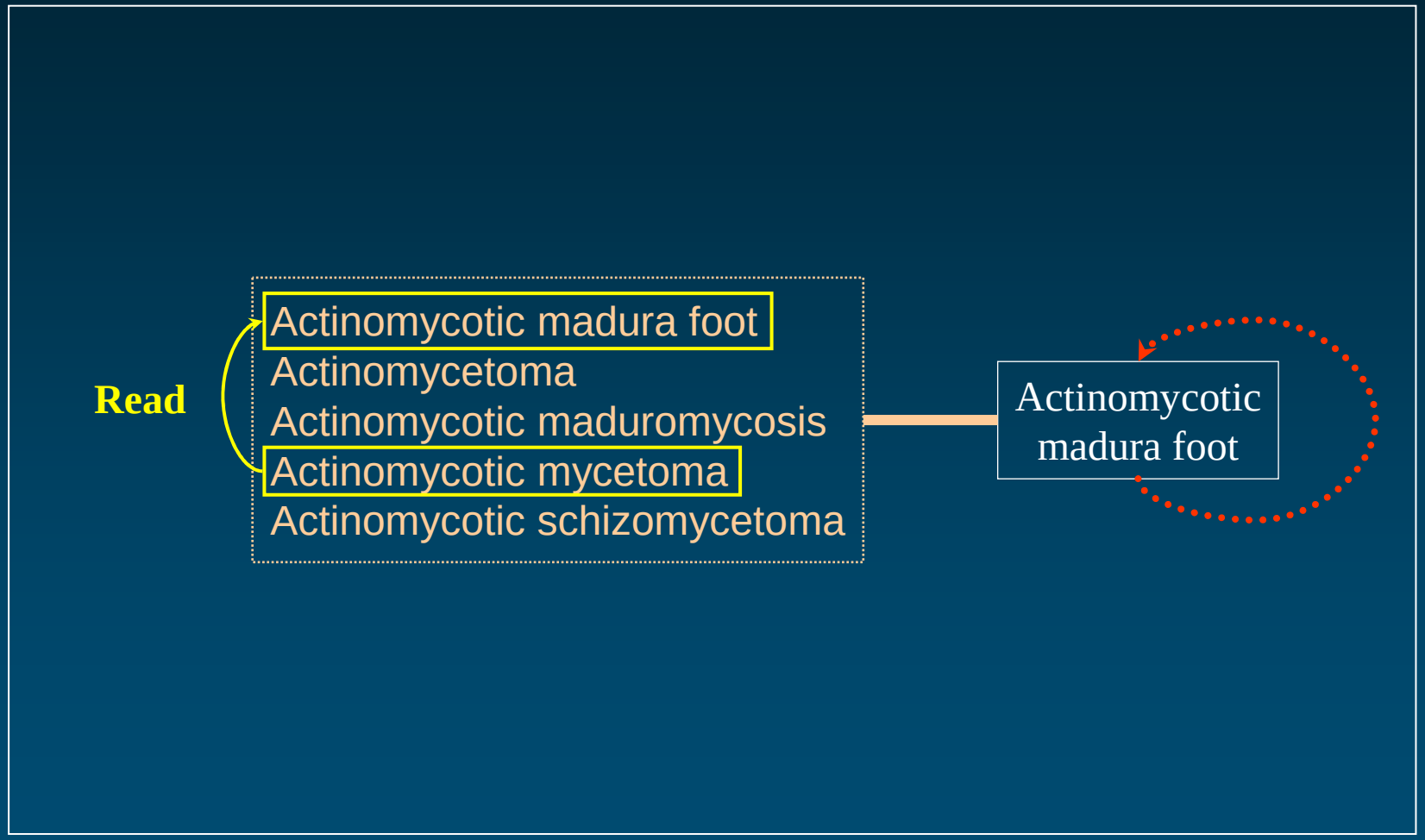
  - Reflexive: CUI1 = CUI2
  - Direct: CUI1|PAR/RB|CUI2 and CUI1|CHD/RN|CUI2
  - Indirect: graph analysis (depth-first search)
- Break cycles
  - Reflexive: remove all (or ignore)
  - Direct: remove (at least) one of the two links
    - Contexts (original trees), redundancy
  - Indirect: remove (at least) one link
    - Manual review

Example of use

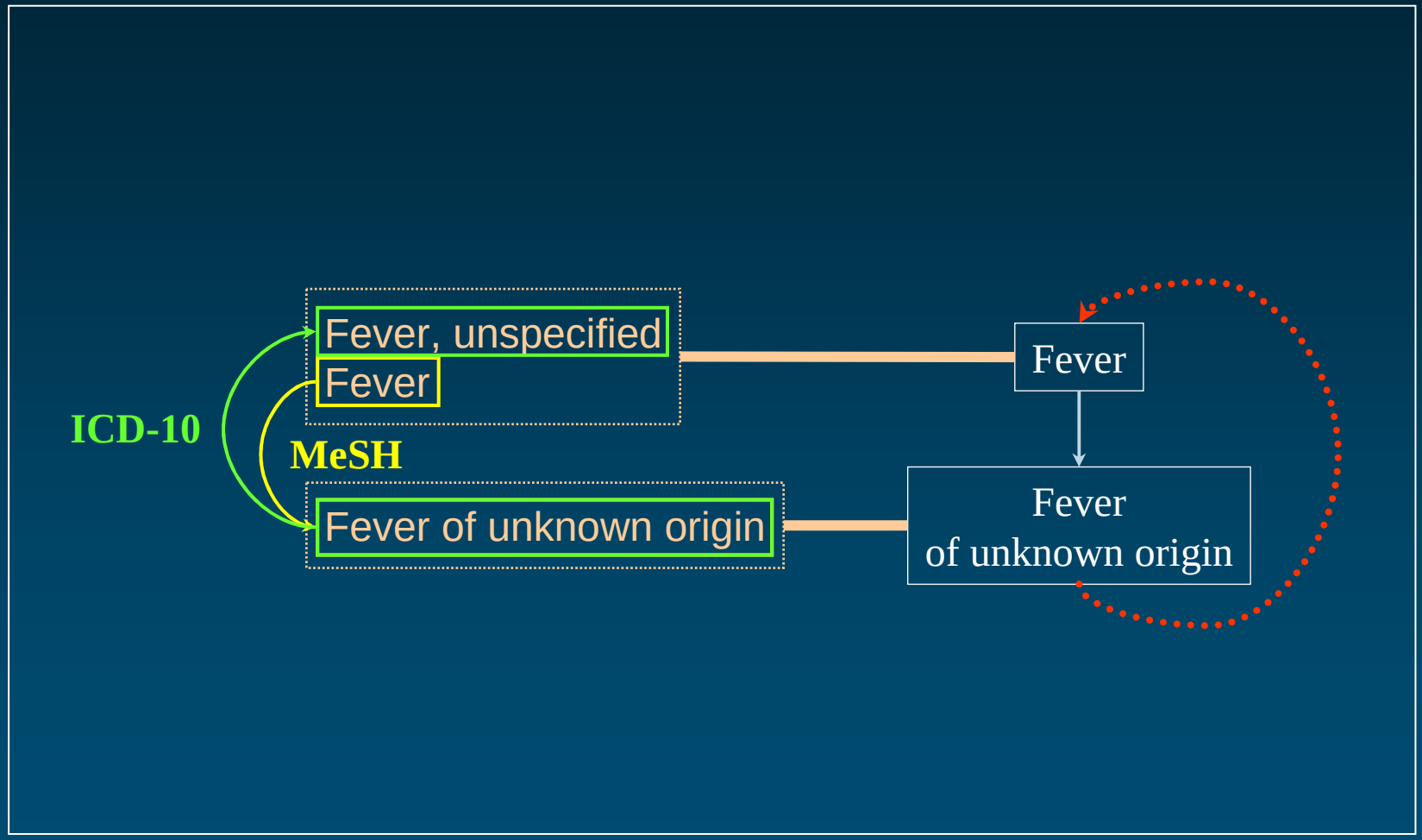
- Create an acyclic Metathesaurus
- Removed
  - 13,000 reflexive relationships
  - 1800 direct relationships
  - 120 indirect relationships

Bodenreider O.
Circular Hierarchical Relationships in the UMLS: Etiology, Diagnosis, Treatment, Complications and Prevention.
Proc AMIA Fall Symp. 2001 (in press) [S78 - Wednesday 8:30am]

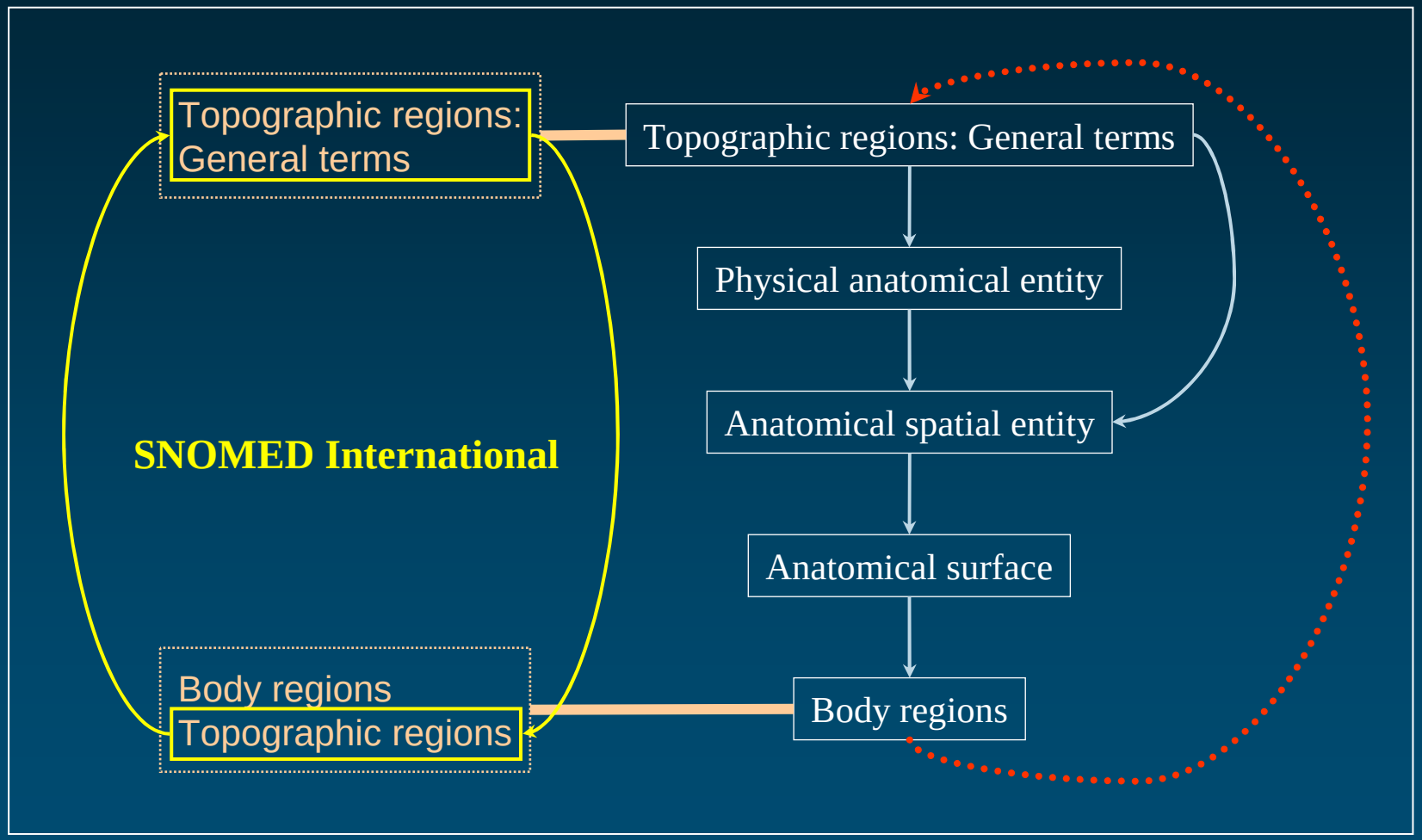
Example Reflexive relationship



Example Direct relationship



Example Indirect relationship



Discussion

- Small number of cycles, but large number of concepts having at least one cycle among the graph of their ancestors / descendants
- Methods based on redundancy
  - are no substitute for a careful review
  - But represent a trade-off between cost and efficacy
- Controls based on structure could be performed at the level of data entry

Customize Relationships

❸ Statistical Approach

Background Statistical Knowledge

- Several kinds of knowledge in the Metathesaurus recorded as interconcept relationships
  - Symbolic: based on the meaning (MRREL)
    - "Addison's disease" isa "disease"
    - "Addison's disease" associated with "Addisonian crisis"
  - Statistical: based on the co-occurrence of MeSH descriptors in MEDLINE citations (MRCOC)
    - "Addison's disease" coc "adrenal glands" [19/808]
    - "Addison's disease" coc "prostatic neoplasms" [2/808]
    - "Addison's disease" coc "quality of life" [2/808]

An example from MEDLINE

Cugini P, Letizia C, Cerci S, Di Palma L, Battisti P, Coppola A, Scavo D.
A chronobiological approach to circulating levels of renin, angiotensin-converting enzyme, aldosterone, ACTH, and cortisol in Addison's disease.
***Chronobiol Int* 1993 Apr;10(2):119-22**

This study deals with a chronobiological approach to the circadian rhythm of the renin-angiotensin-aldosterone system (RAAS) and the ACTH-cortisol axis (ACA) in patients with Addison's disease (PAD). The aim is to explore the mechanism(s) for which the circadian rhythmicity of the RAAS and ACA takes place. The study has shown that both the RAAS and ACA are devoid of a circadian rhythm in PAD. The lack of rhythmicity for renin and ACTH provides indirect evidence that their rhythmic secretion is in some way related to the circadian oscillation of aldosterone and cortisol. This implies a new concept: a positive feedback may be included among the mechanisms which chronoregulate the RAAS and ACA.

PMID: 8388783, UI: 93272348

- Addison's Disease/physiopathology
- Addison's Disease/blood*
- Adolescence
- Adult
- Aldosterone/blood*
- Circadian Rhythm*
- Corticotropin/blood*
- Female
- Human
- Hydrocortisone/blood*
- Male
- Middle Age
- Peptidyl-Dipeptidase A/blood*
- Renin/blood*

Background Co-occurences

- Relationships

  - Pair of concept identifiers
  - Frequency of co-occurrence
  - Source of co-occurrence
- Semantics of the relationship: undefined
  - Some redundancy with symbolic relationships
  - “Addison’s disease” coc “prostatic neoplasms” [2/808]

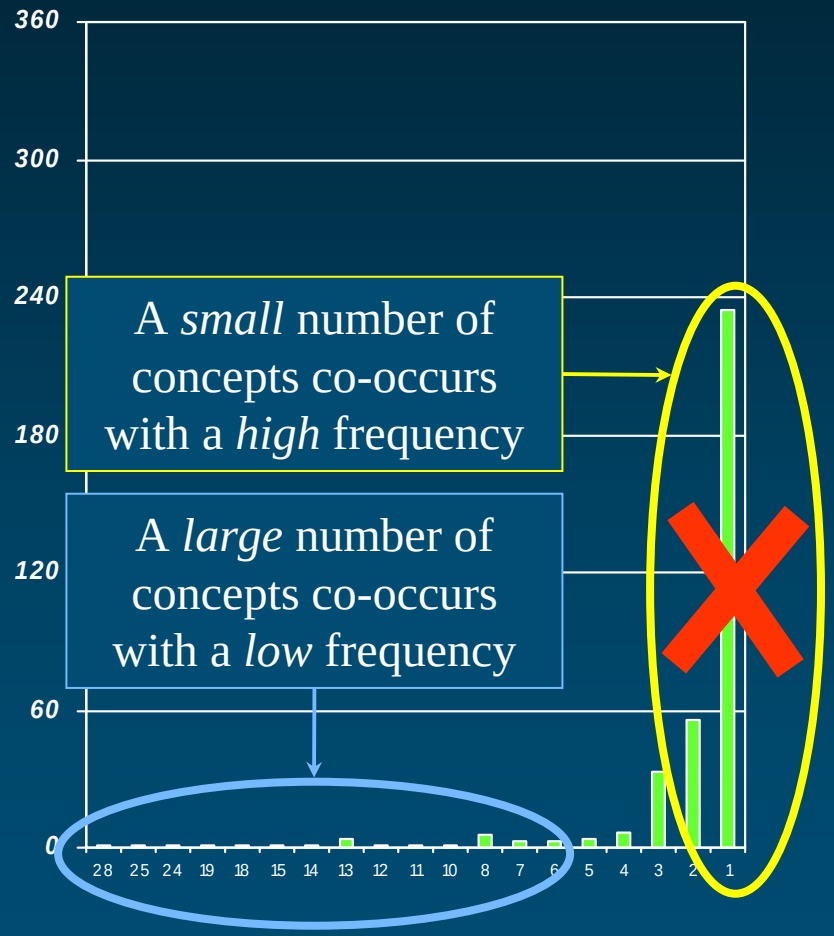
> - *Addison's disease secondary to prostatic carcinoma. A case report.*
> - *Retropubic radical prostatectomy in a patient with chronic adrenal insufficiency*

Background Co-occurences

- Only co-occurrence between “starred” descriptors is recorded in the Metathesaurus
- Relative frequency of co-occurrence
  - Freq(A and B) / Freq(A)
  - Freq(A and B) / Freq(B)
  - Surrogate for the strength of the link
- Frequency distribution may help select the most significant co-occurrences

Addison's Disease: Co-occurring concepts

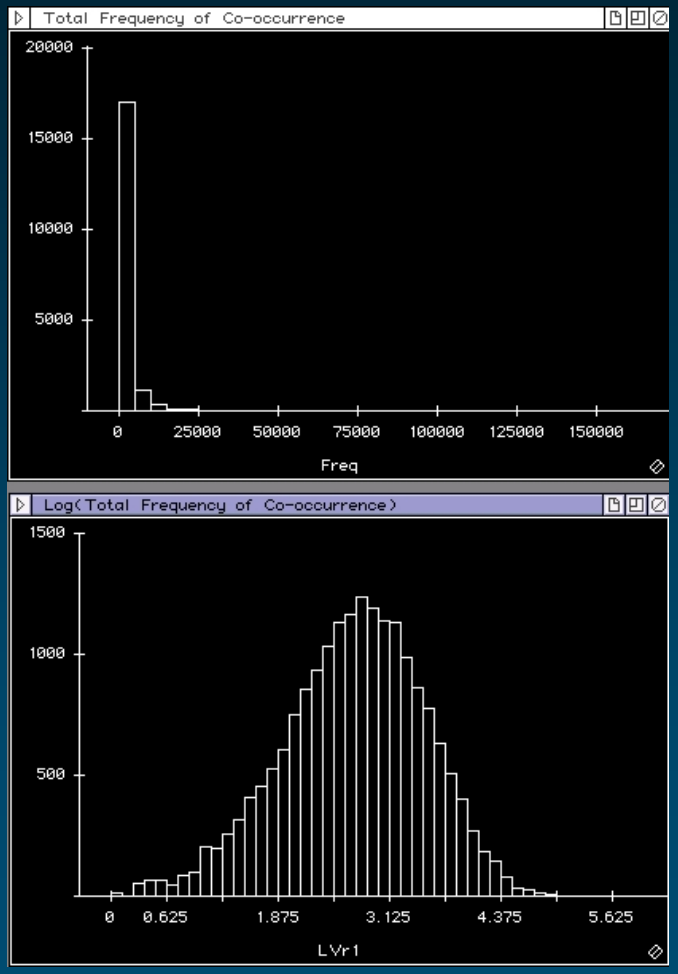
```
28  Autoimmune Diseases
25  Autoantibodies
24  Hydrocortisone
19  Adrenal Glands
18  Steroid 21-Monooxygenase
15  Corticotropin
14  Adrenal Gland Neoplasms
13  Adrenal Cortex
13  Adrenal Gland Diseases
13  Glucocorticoids
13  Polyendocrinopathies, Autoimmune
12  Diabetes Mellitus, Insulin-Dependent
11  Tuberculosis, Endocrine
10  Adrenoleukodystrophy
 8  Adrenal gland hypofunction
 8  Autoantigens
 8  Cushing Syndrome
 8  Hypothyroidism
 8  Tuberculosis
 8  Chronic lymphocytic thyroiditis
    […]
 1  Circadian Rhythm
    […]
```



Total frequency of co-occurrence

- Number of co-occurring concepts
  - Min: 1
  - Max: 164,762
  - Median: 585

164762	Brain
137102	Liver
126009	Neurons
105382	Calcium
102109	Postoperative Complications
101955	DNA-Binding Proteins
93425	Breast Neoplasms
86878	RNA, Messenger
83578	Transcription Factors
82987	Escherichia coli
82840	T-Lymphocytes
82629	Aging
81442	Hypertension



Motivation

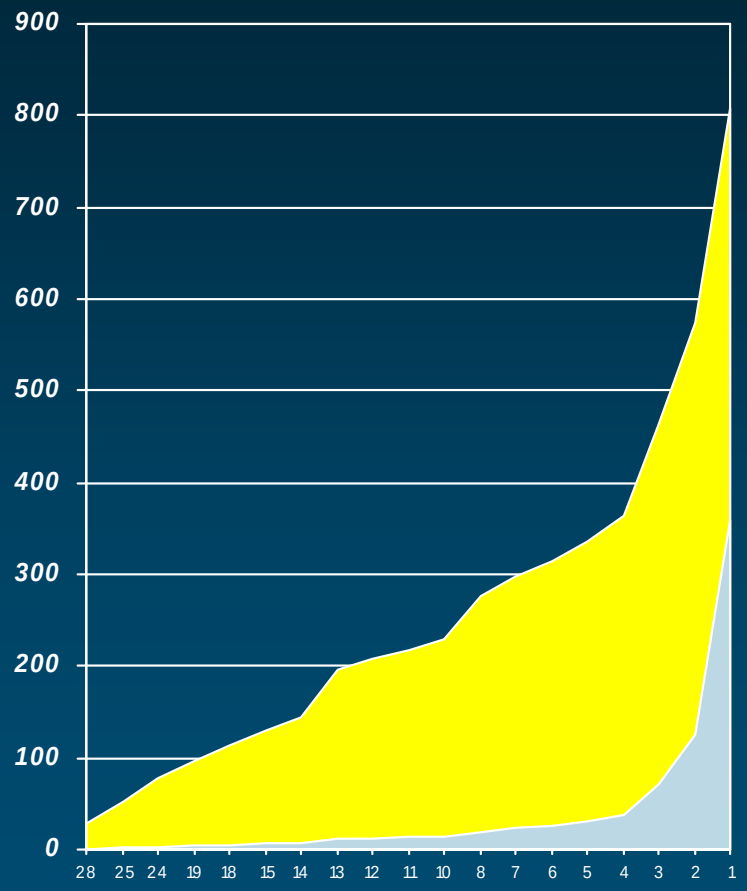
- Reduce the volume
- Select significant associations
  - For display purposes
  - Discover unexpected associations
  - Select candidate associative relationships for UMLS editors to review

Methods

- Threshold on relative frequency of co-occurrence
  - Fixed threshold
    - Absolute (e.g., at least 2)
    - Relative (e.g., at least 1%)
  - Percentile
    - e.g., 90th percentile
    - Problem with long distribution tails
  - Dynamic approach
    - Smallest number of pairs representing the largest fraction of the total frequency

Methods

- 19 classes (concepts with the same frequency)
- Total frequency: 808
- Add classes until the benefit of adding the next class becomes insignificant



Example of use Visualization

- Display only a reasonable number of co-occurring concepts
- Addison's disease
  - Co-occurring concepts: 360
  - *Displayed:* *126 (35%)*
  - Total frequency of co-occurrence: 808
  - *Represented:* *574 (71%)*

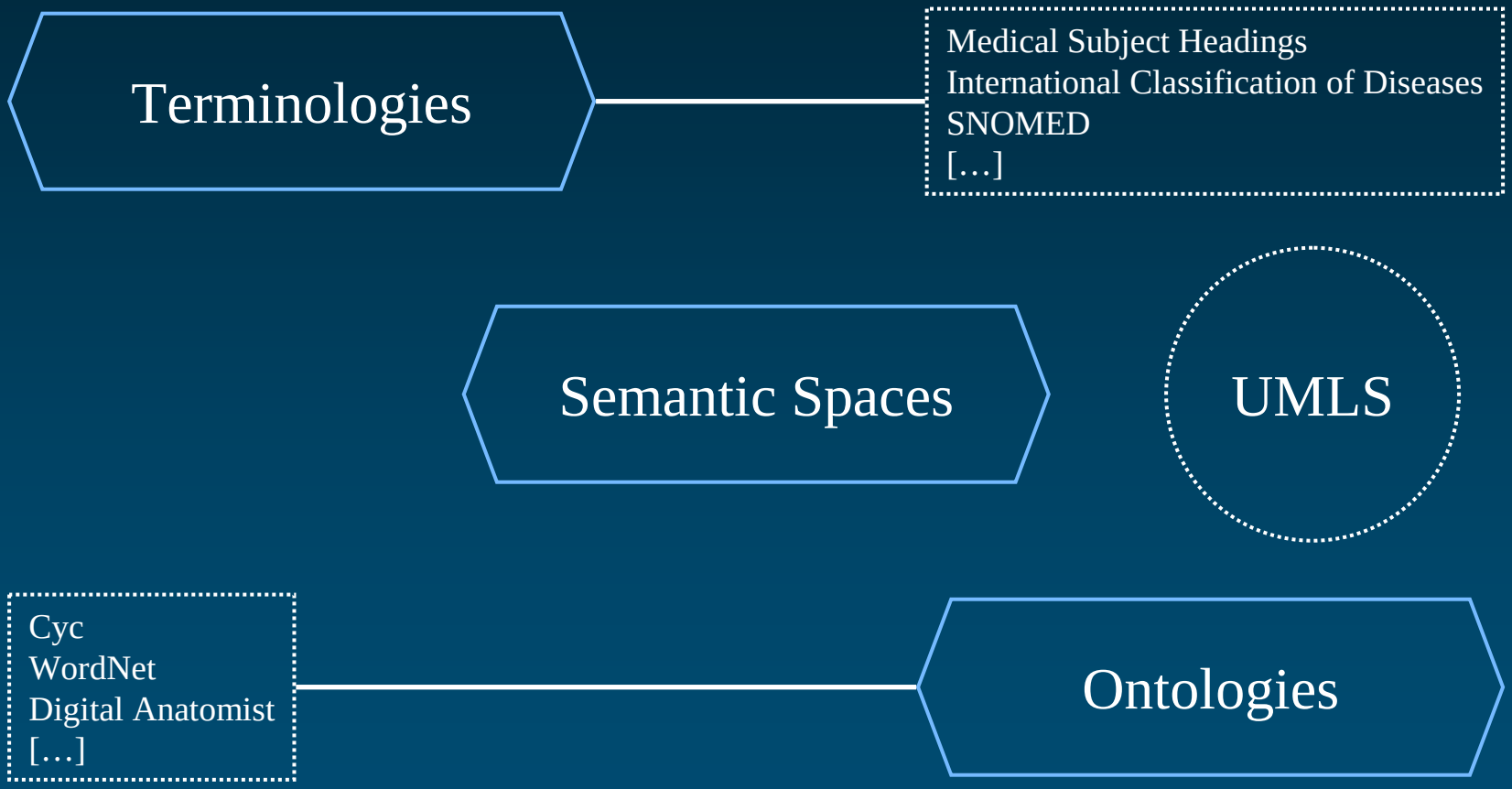
Discussion

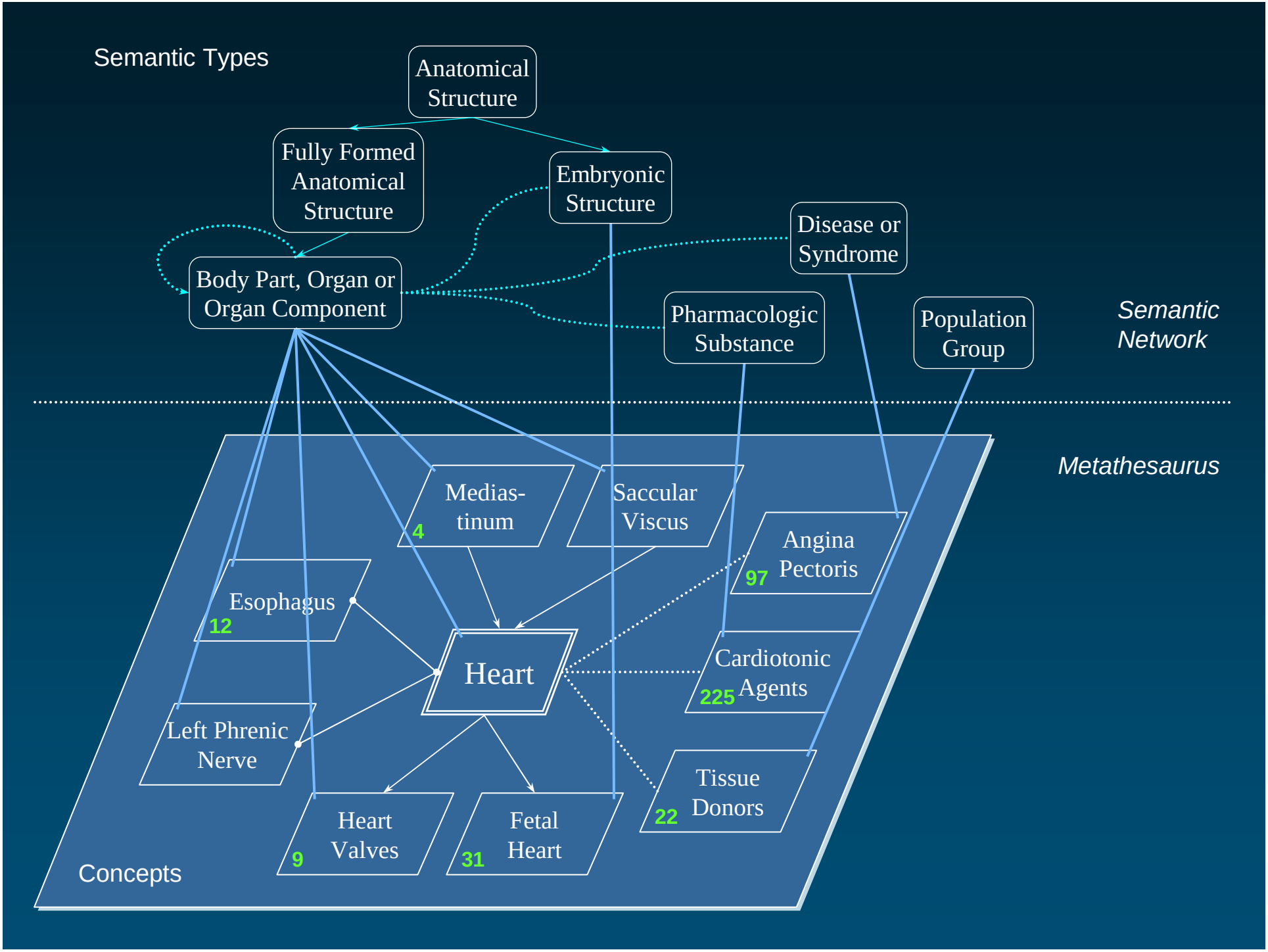
- Only 6 percent of the relationships between co-occurring concepts are redundant with symbolic relationships in the Metathesaurus
- A more sophisticated statistical analysis is necessary to refine the filter
- Additional filters may be applied
  - E.g., minimum value for the total frequency of co-occurrence

Outline of Tutorial

- Why customize? — Betsy Humphreys
- Metathesaurus basics — Olivier Bodenreider
- How to customize?
  - Customize sources (MetamorphoSys) — L. Roth & S. Srinivasan
  - Customize strings — Olivier Bodenreider
  - Customize synonyms
  - Customize relationships
  - Customize concept spaces
- Adding “local” terminology — Bill Hole

Background Knowledge organization



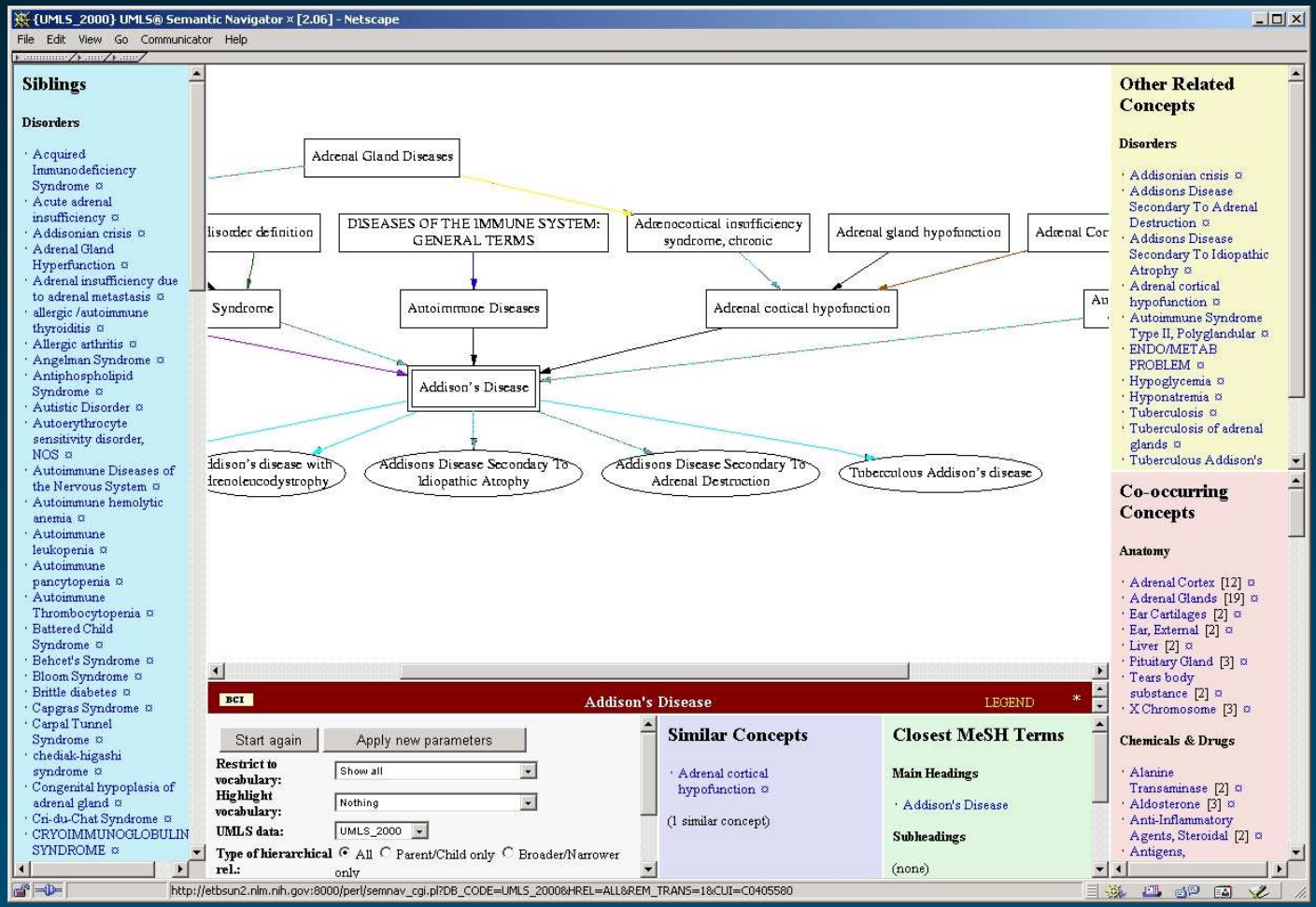

Semantic Types
Anatomical Structure
Fully Formed Anatomical Structure
Embryonic Structure
Disease or Syndrome
Body Part, Organ or Organ Component
Pharmacologic Substance
Population Group
Semantic Network
Metathesaurus
Medias-tinum
4
Saccular Viscus
Angina Pectoris
97
Esophagus
12
Heart
Cardiotonic Agents
225
Left Phrenic Nerve
Tissue Donors
22
Heart Valves
9
Fetal Heart
31
Concepts

UMLS Semantic Navigator

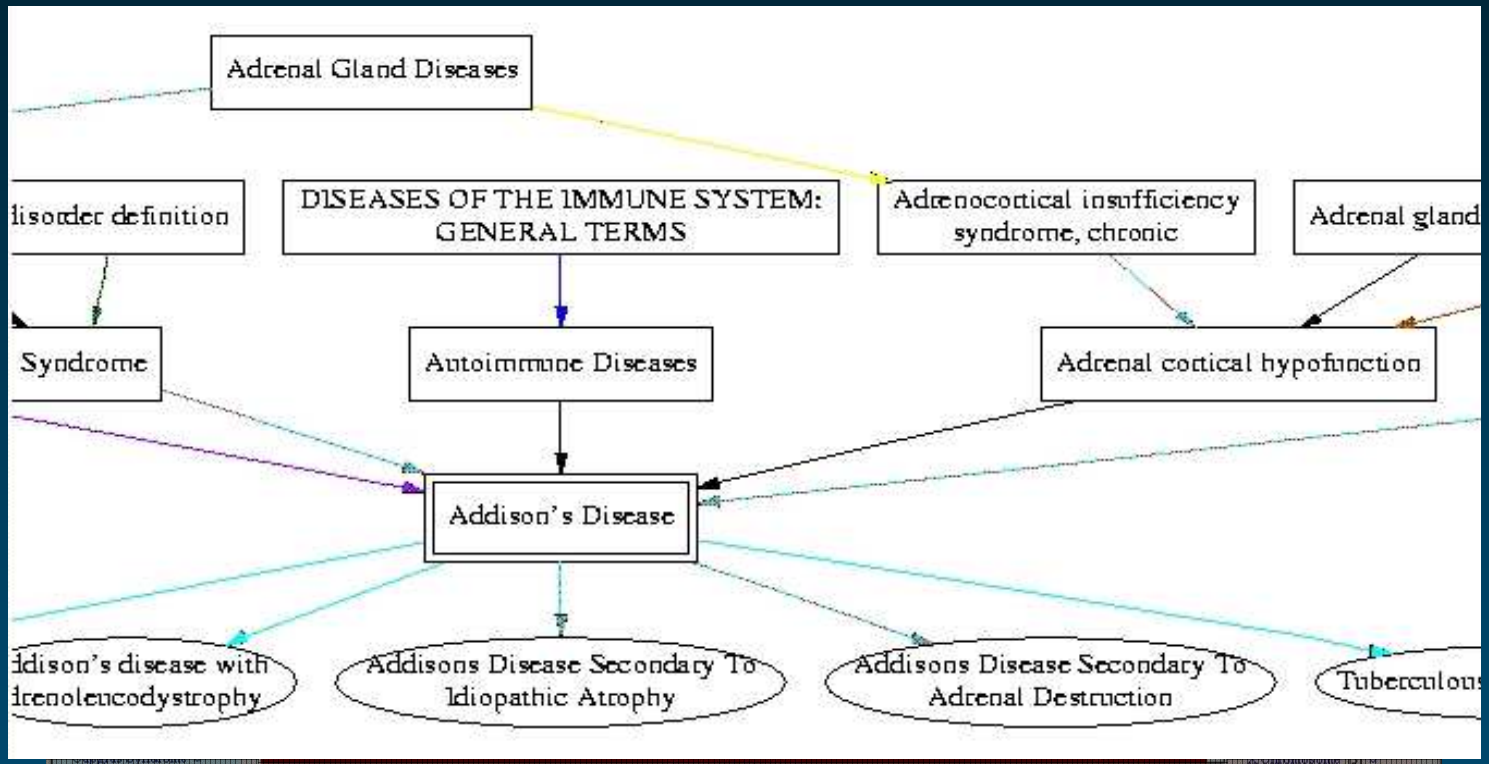
- Features
  - All relationships presented simultaneously
    - Metathesaurus relationships
    - Semantic network relationships
  - Hierarchical relationships presented graphically
  - Dynamic and navigable

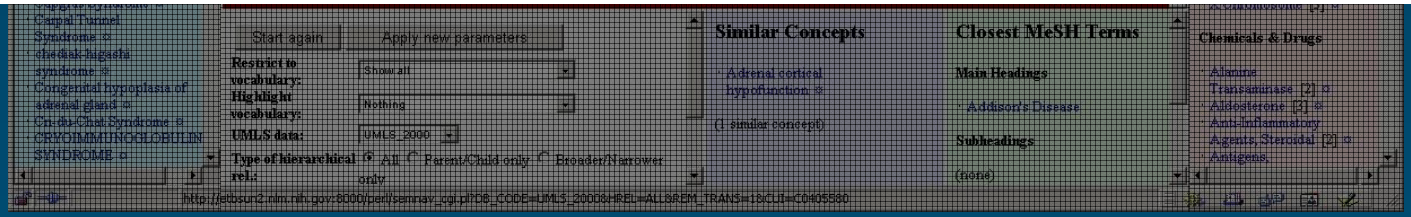
umlsks.nlm.nih.gov → Resources → Semantic Navigator

UMLS Semantic Navigator



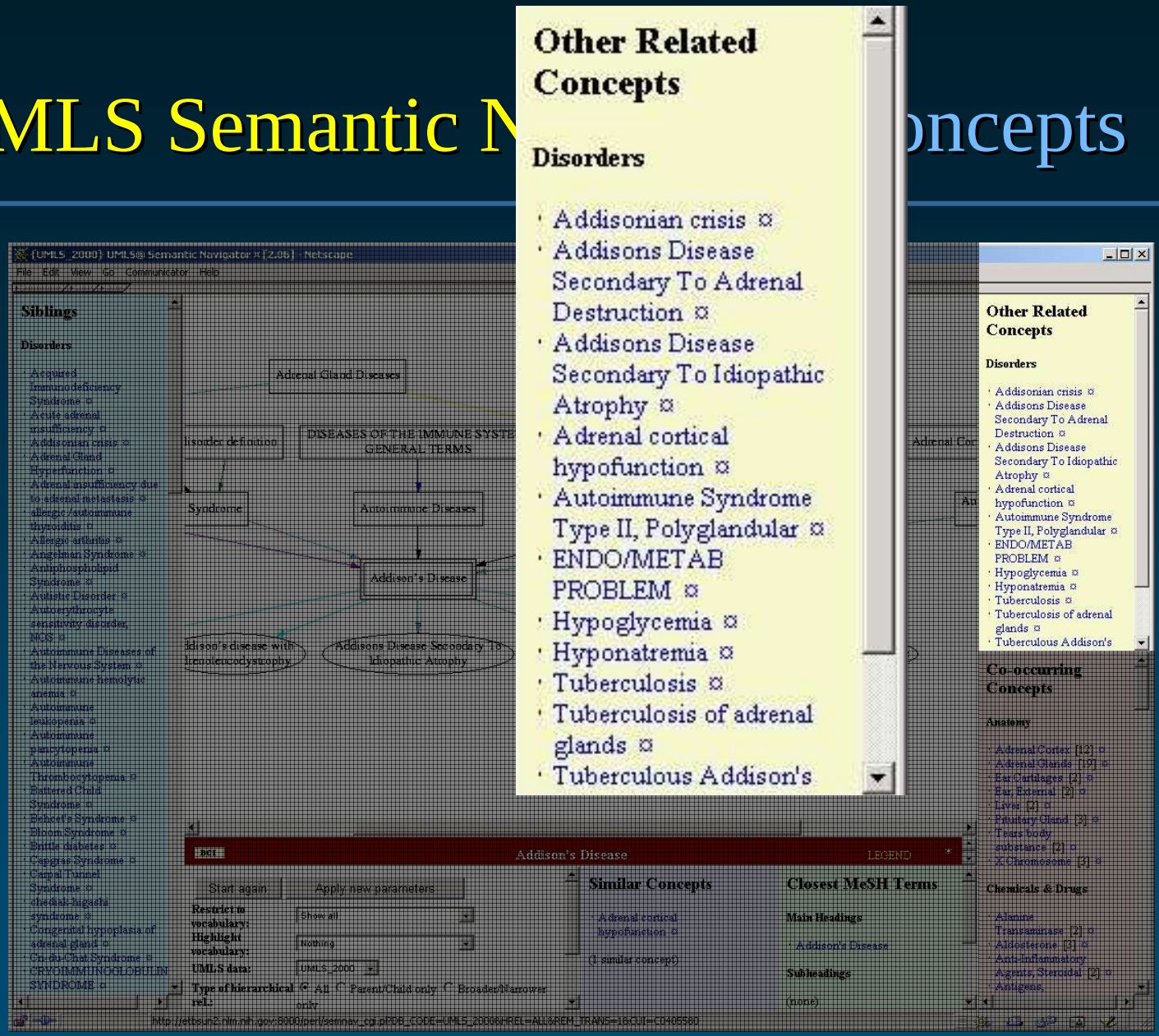
UMLS Semantic Navigator Concepts



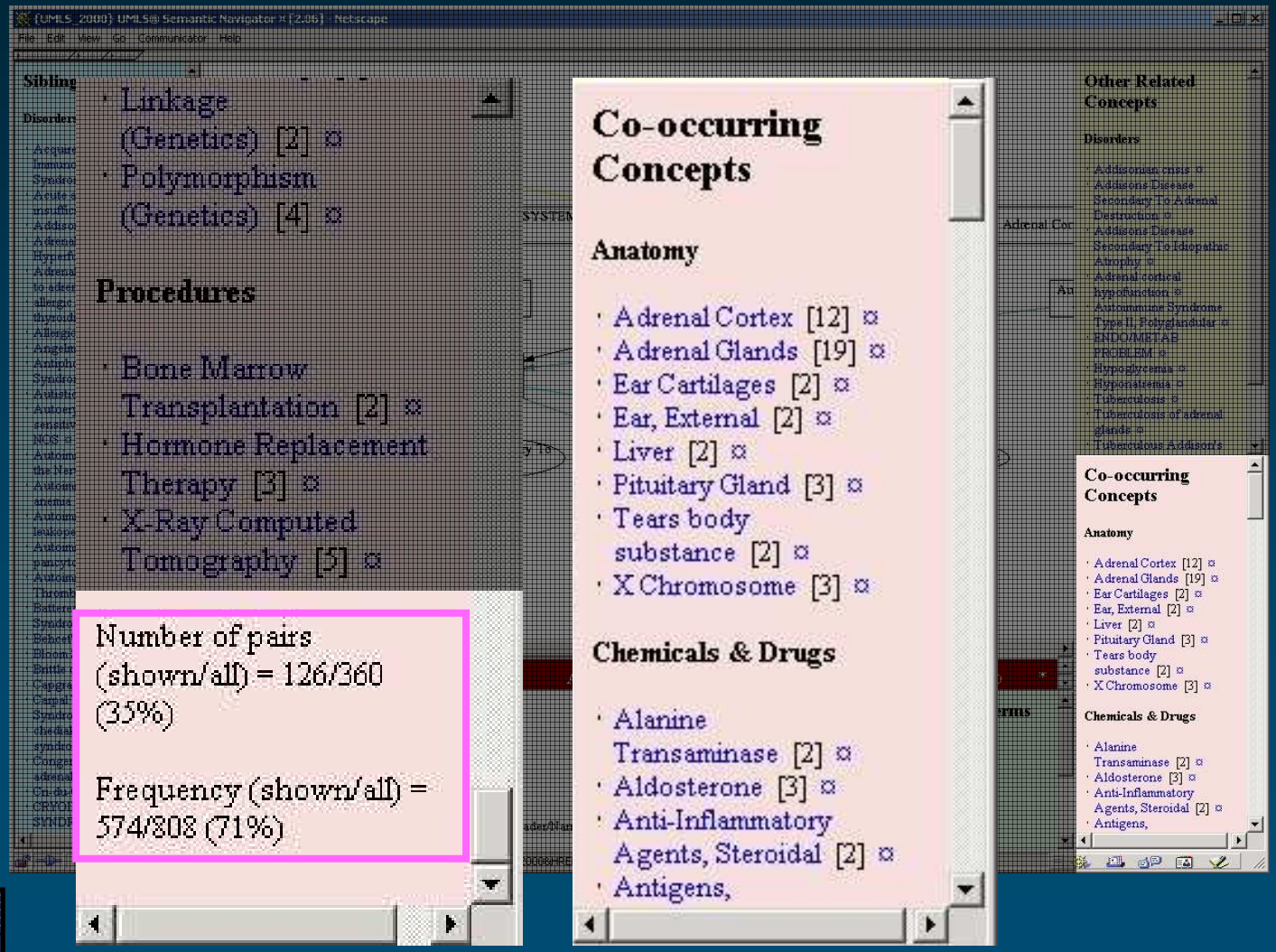


UMLS Se igator Concepts

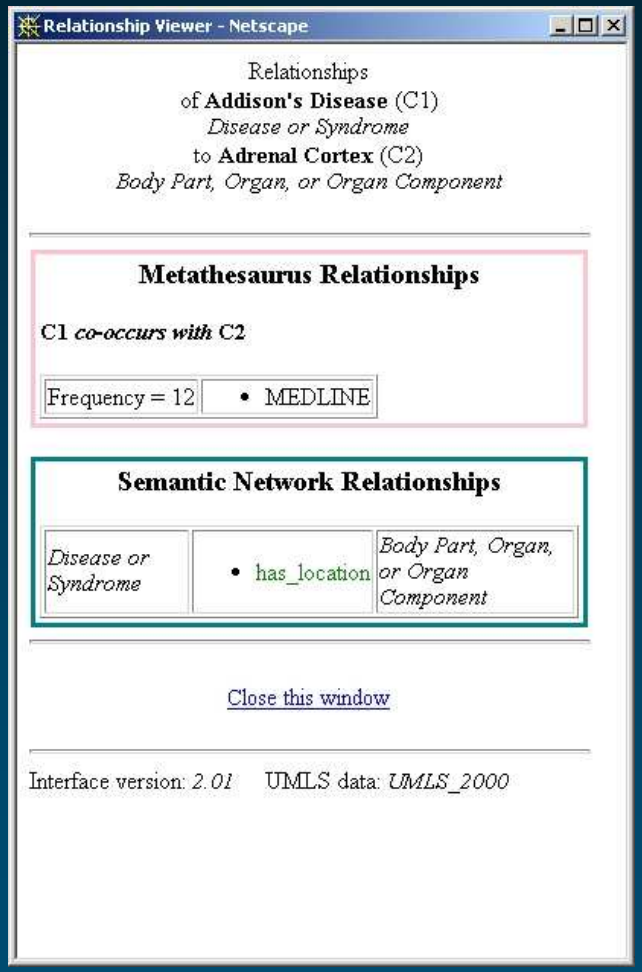
UMLS Semantic N oncepts

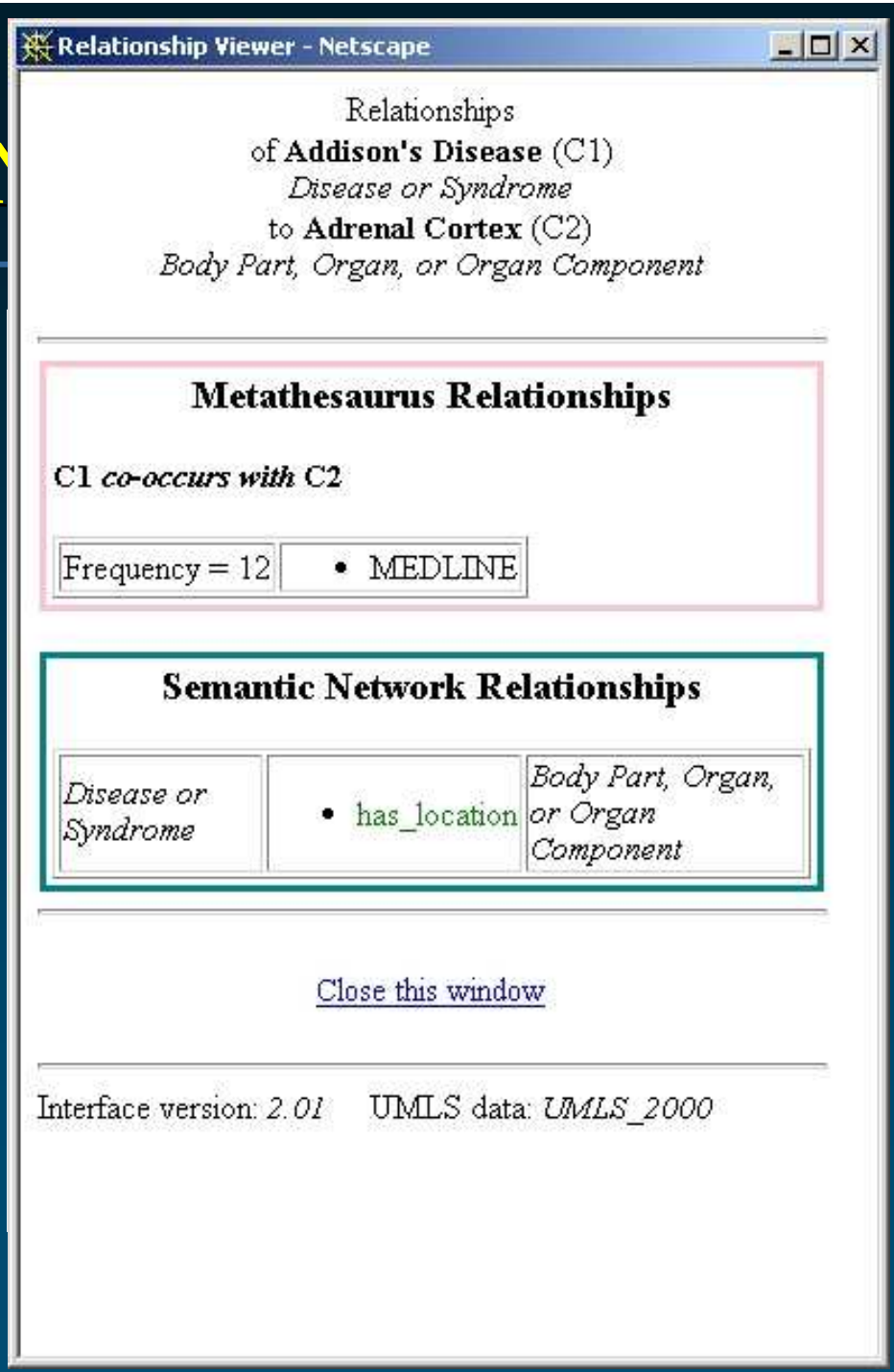


UMLS Semantic Navigator Concepts



UMLS Semantic N





NATIONAL LIBRARY OF MEDICINE

Motivation

- Reduce volume
  - Concepts
  - Relationships
  - Both
- Reduce ambiguity

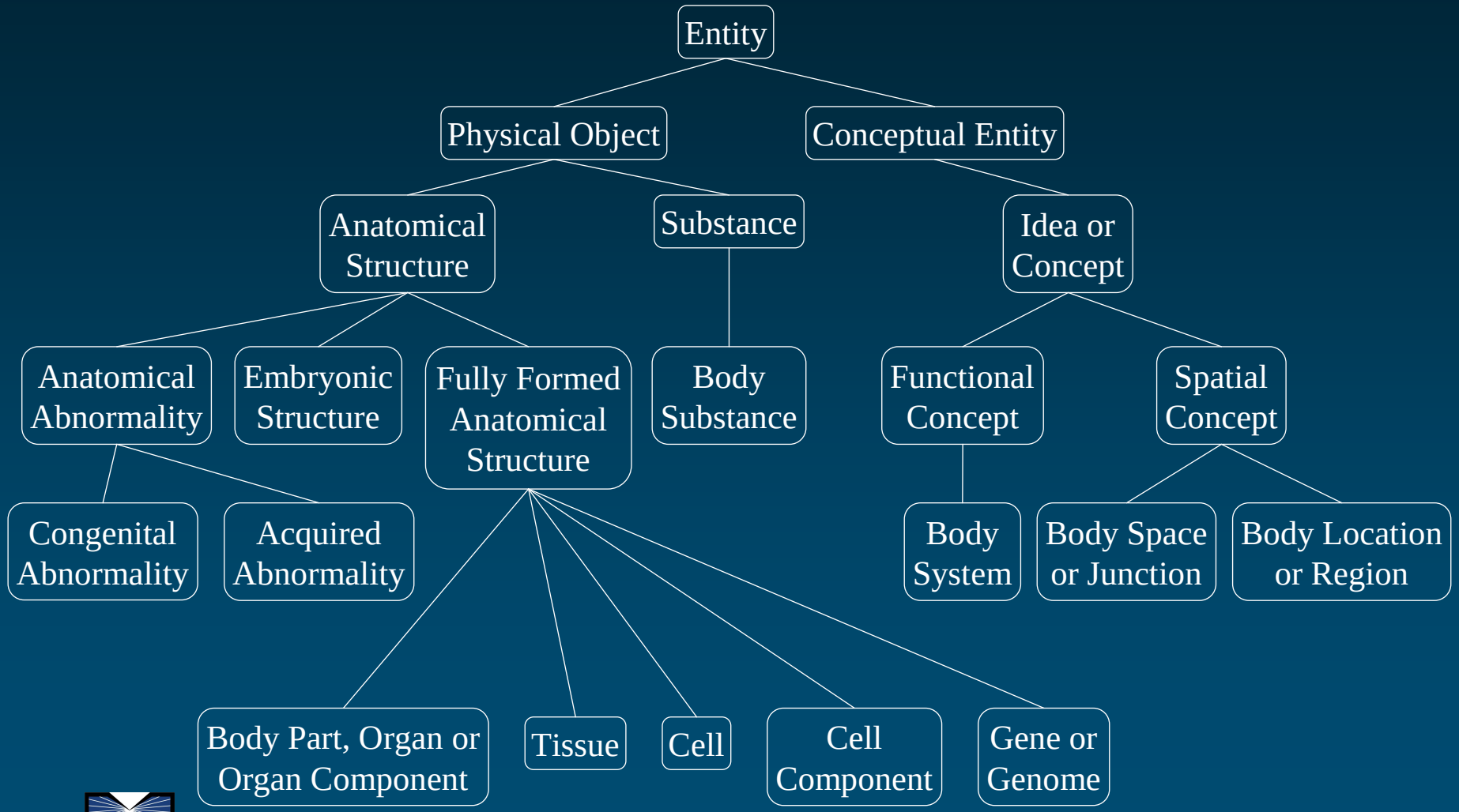
Methods

- Based on the categorization: Semantic groups

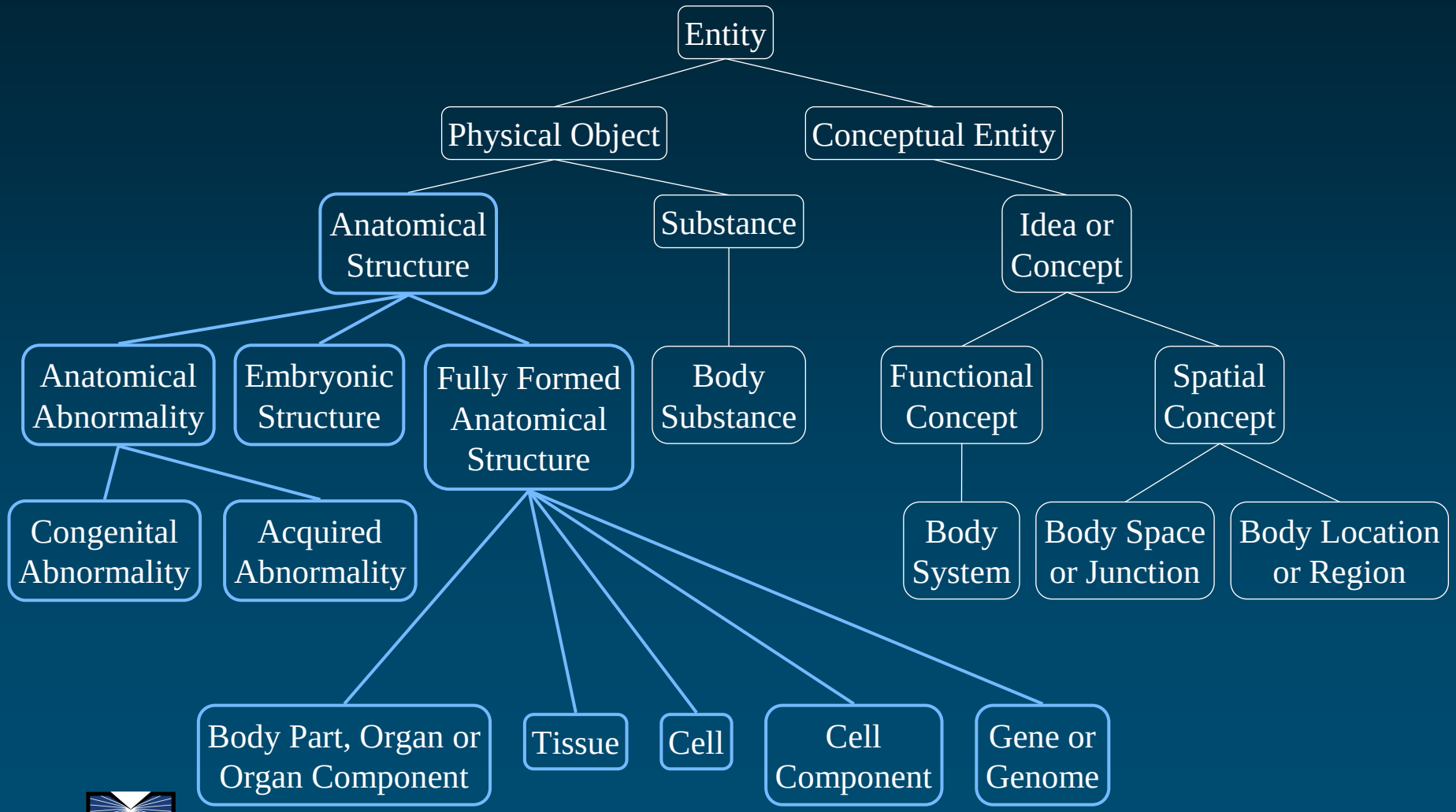
> McCray A.T, Burgun A., Bodenreider O.
> Aggregating UMLS semantic types for reducing conceptual complexity.
> Medinfo 2001;10 Pt 1:216-220.

- Based on inter-concept relationships:
  - Transitive reduction (structural)
  - Semantic distance (symbolic + statistical)

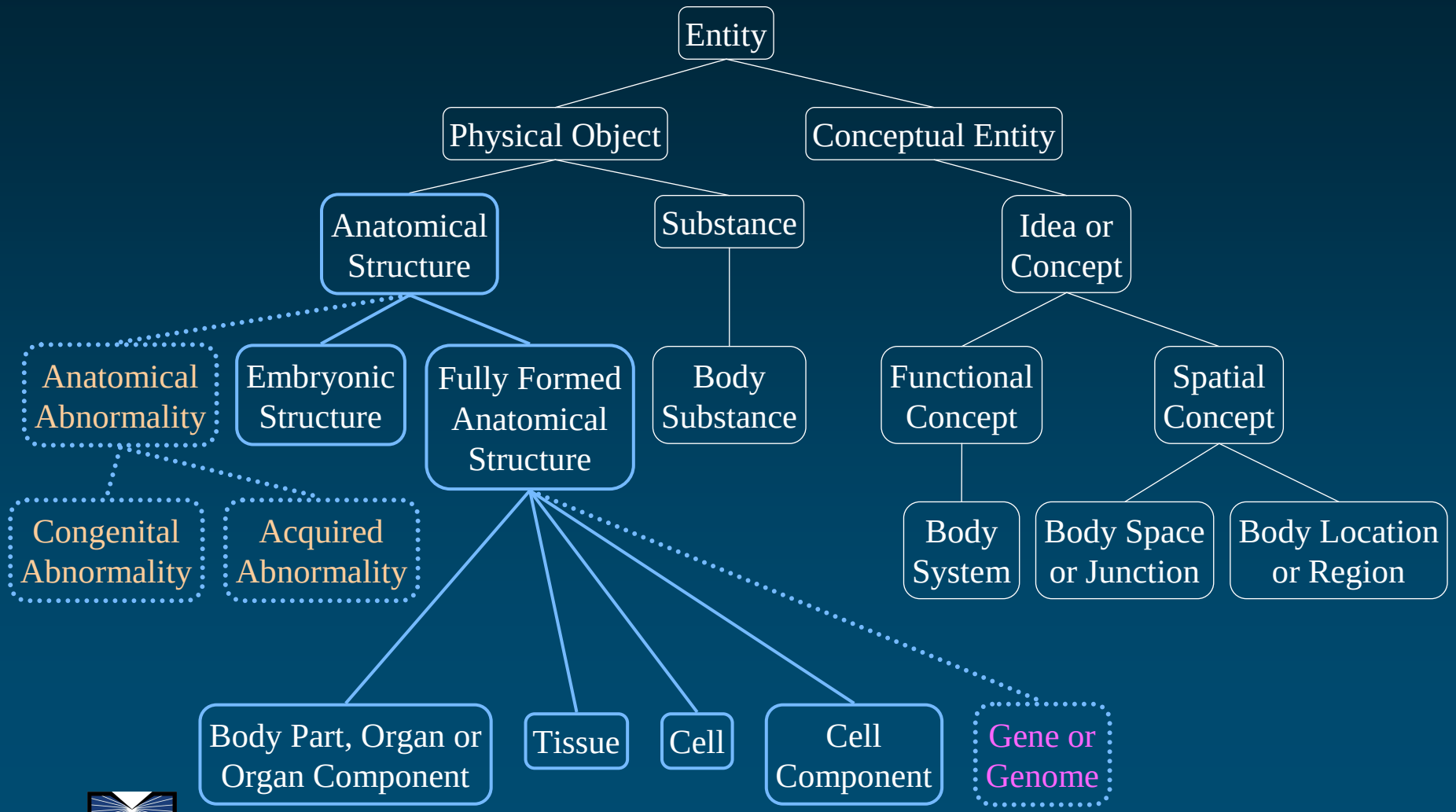
Semantic Network



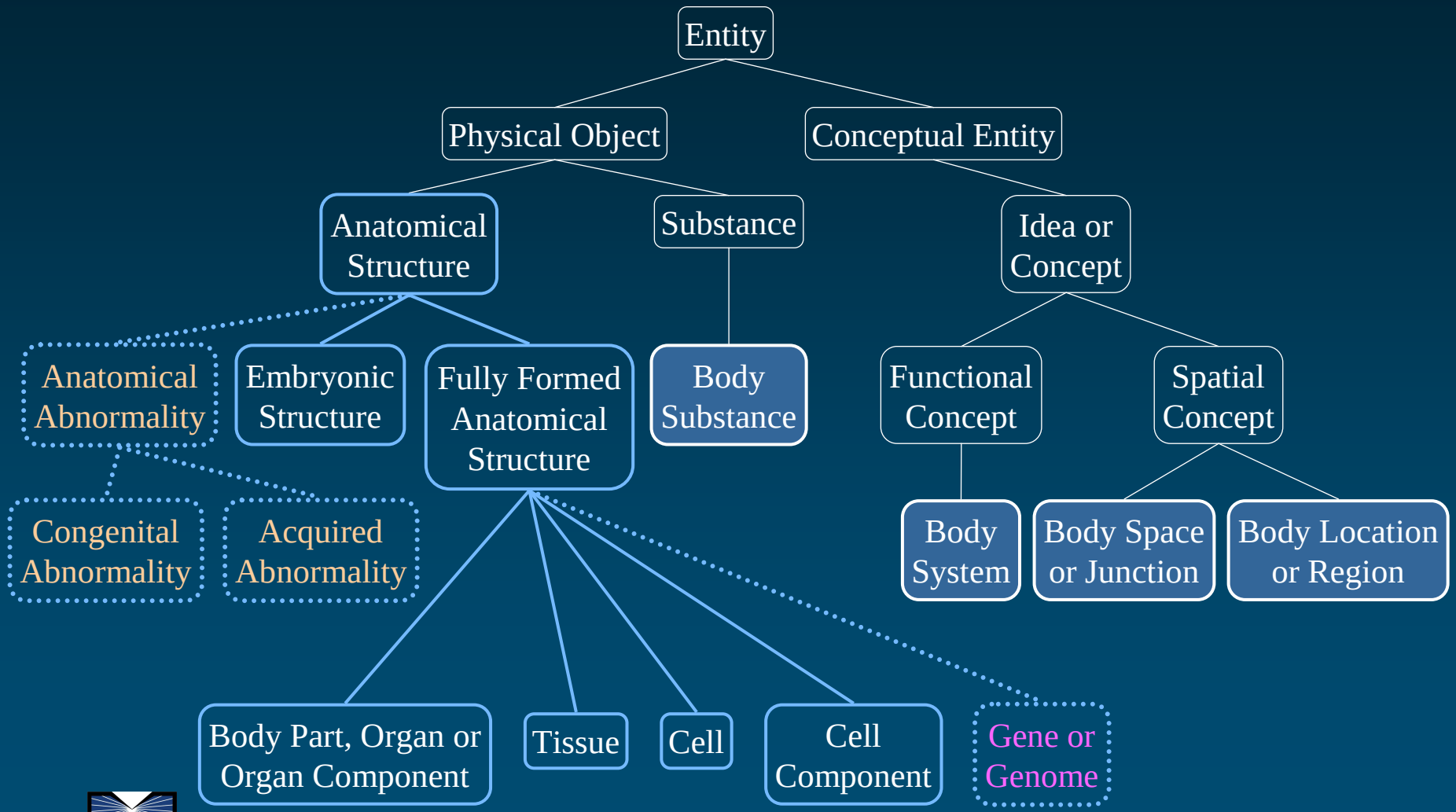
Semantic Network Anatomy subtype



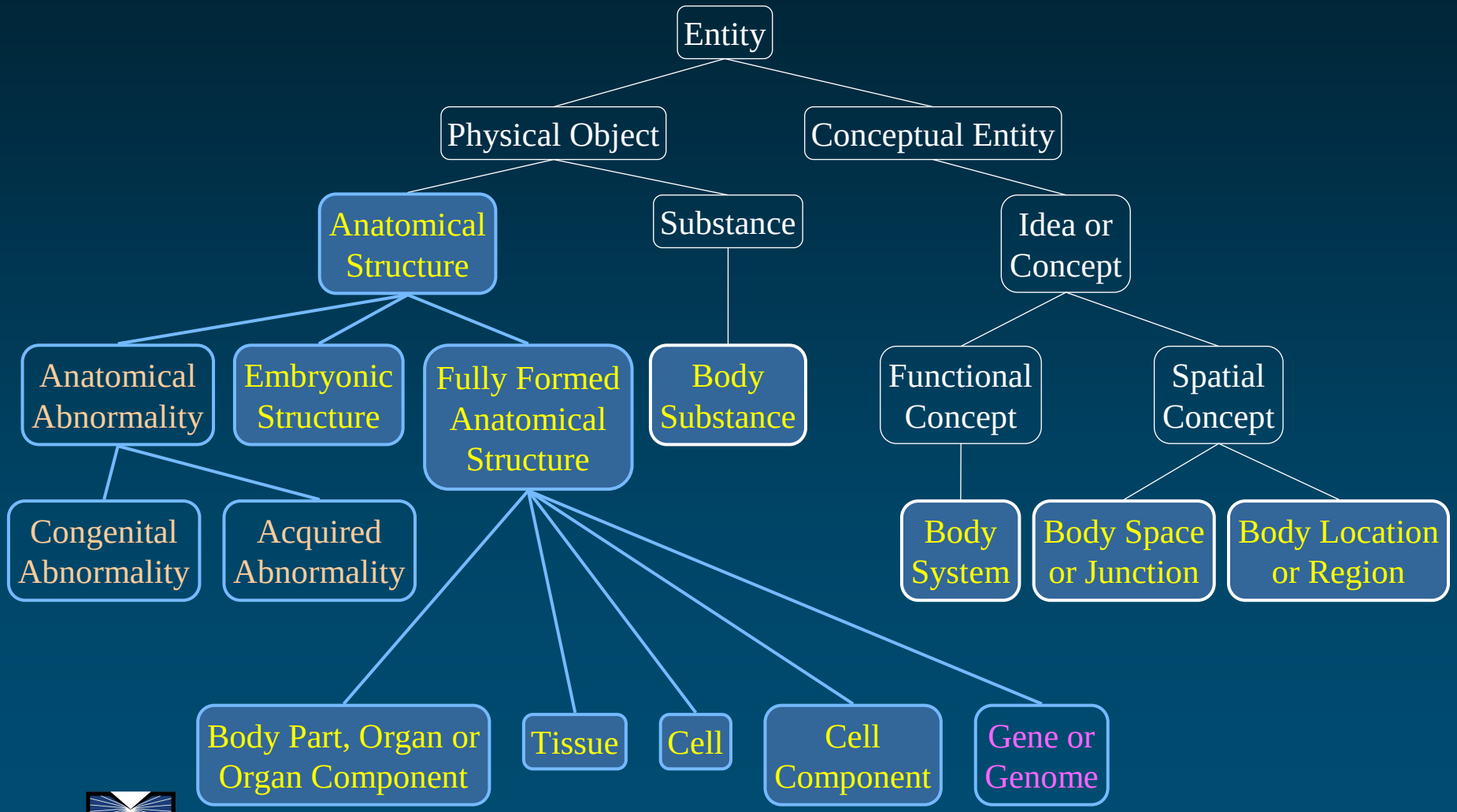
Semantic Network Detach some types



Semantic Network Attach some types



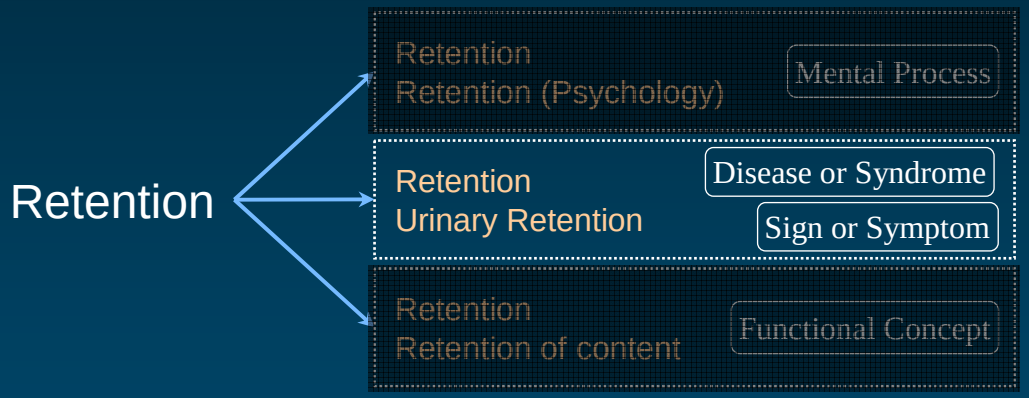
Semantic Network SG Anatomy



Example of use

- Disambiguate
- Extract semantic subspaces
  - Major semantic axis (e.g., anatomy)
  - Body system (e.g., cardiology)
  - Procedure (e.g., transplantation)
- Simplify representation for visualization purposes

Example of use Disambiguate

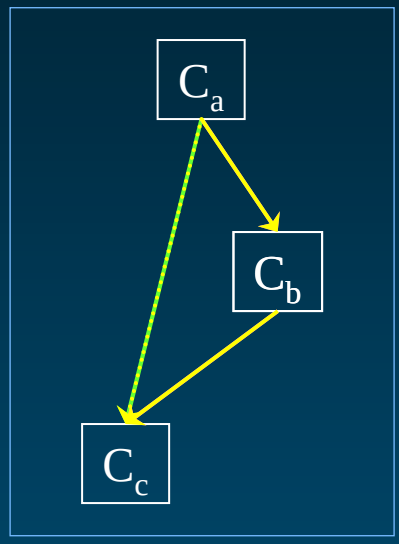


Example of use Semantic subspaces

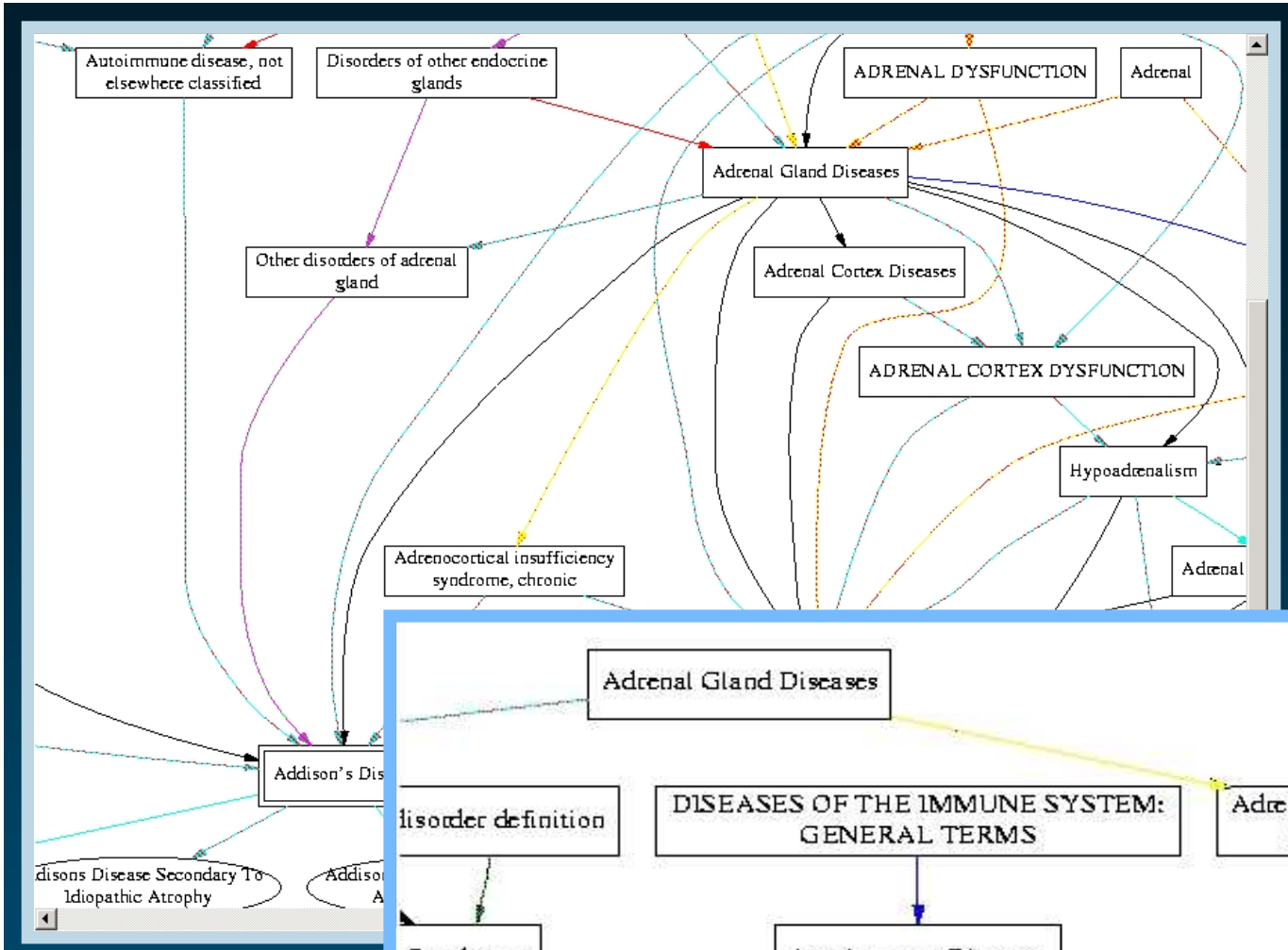
- Major semantic axis (e.g., anatomy)
  - Use semantic groups
- Body system (e.g., cardiology)
  - Use interconcept relationships
  - Combine relationships: Family
    - Uncles = siblings of parents
    - Cousins = children of uncles
- Procedure (e.g., transplantation)

Example of use Simplify representation

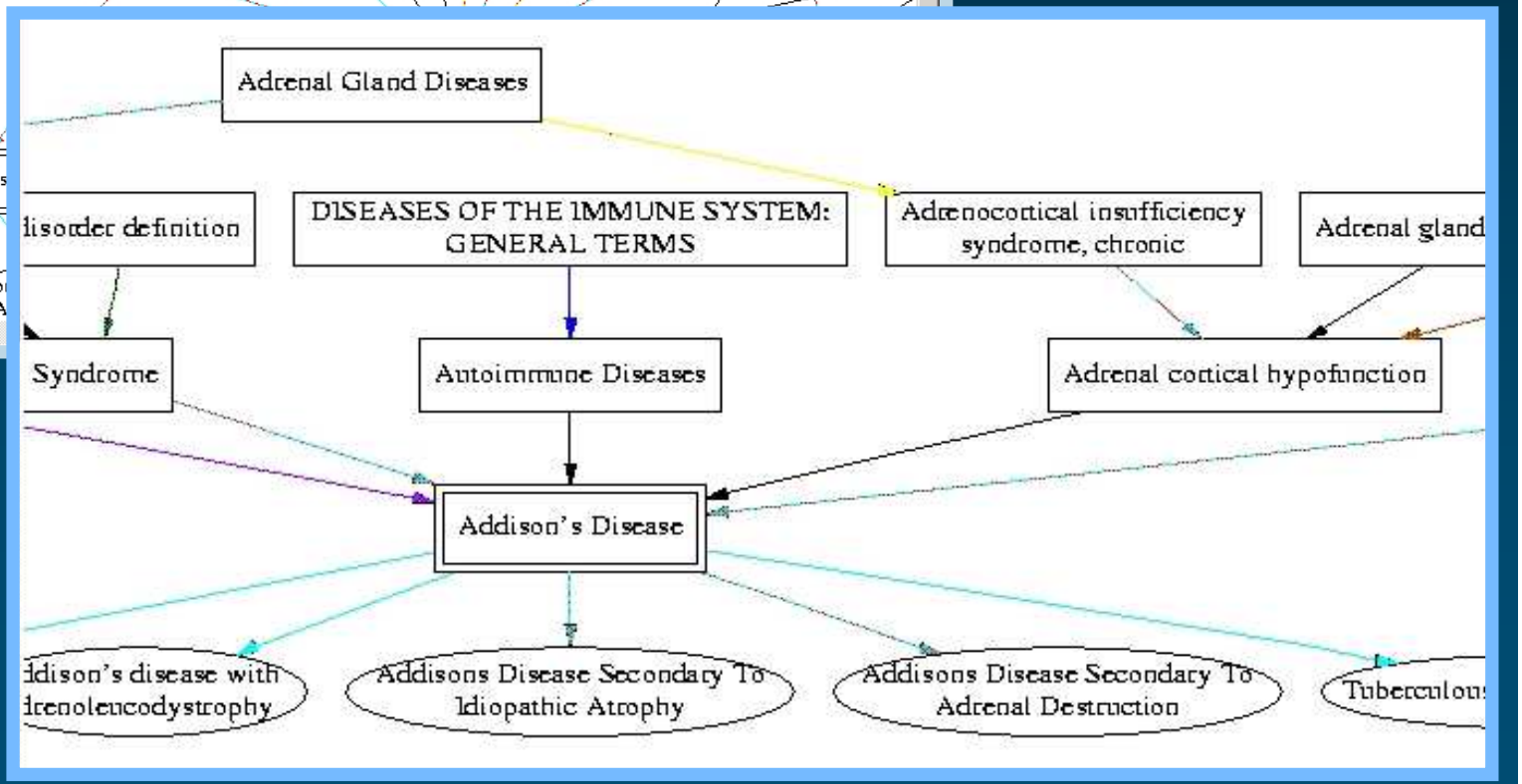
- Hide “redundant” relationships
- Structural approach
- Transitive reduction



All relationships



Transitive reduction



Discussion

- Alternative approaches
  - Core concepts
  - Concepts found in multiple sources
- Semantic distance
  - Work in progress

Outline of Tutorial

- Why customize? — Betsy Humphreys
- Metathesaurus basics — Olivier Bodenreider
- How to customize?
  - Customize sources (MetamorphoSys) — L. Roth & S. Srinivasan
  - Customize strings — Olivier Bodenreider
  - Customize synonyms
  - Customize relationships
  - Customize concept spaces
- Adding “local” terminology — Bill Hole

Adding “local” terminology

- Vocabularies not in the UMLS?
- Local terms or terminologies?
- Increments to reference terminologies?

Two key questions

- Are the *meanings* already in the Metathesaurus?
- How will you maintain your system as you and the Metathesaurus add names and meanings?

Create Unique Identifiers for *your* Terminology

- For your concepts, use:

  'CA000001 …' as CUIs instead

  of Meta's 'C0000001 ….' for CUIs

- Similarly, use 'LA000001 ...' for LUIs

  and 'SA000001 …' for SUIs, as needed

- Create a table which can map your UIs to UMLS UIs

e.g., `Your CUI|Meta CUI|`

Which of your terms are Meta Synonyms?

- Use the lvg program to normalize your terms
- look for matches to the Normalized String Index (MRXNS).
- Use other sensible approaches to searching:
  - normalized word searches;
  - explore alternate naming styles and conventions

Hole, W.T, Srinivasan, S.
Discovering Missed Synonymy in a Large Concept-Oriented Metathesaurus.
Proc AMIA Fall Symp. 2000;:354-8

Map your terms to Unique Identifiers

- Use Meta CUIs where synonyms are found
- Use *your* CUIs where no synonyms are found
- Store the map for future use

Bonus Add relationships

- As you look for Meta Synonyms, add *relationships to Meta*
- Assign a REL and RELA to label the particular kinds of relationships you need and will use, e.g. to map or aggregate

Updating to a New Meta Release

- Repeat MetamorphoSys and processing scripts used for the previous release
- Re-use previously found UIs for your terms to map synonyms, etc. to the new Meta
- Check for new Meta Concepts which are synonyms of your terms, not previously in Meta
- Check for any deleted CUIs in MRCUI

```
C0435517|1999|SY|C0435516|
C0361163|1998|DEL||
C0785652|2000|SY|C0775088|
```

Sneak Preview of 2002 changes…

- Metathesaurus changes:

  -MedDRA FDA and international "Medical Dictionary for Regulatory Activities Terminology"

  -VANDF "Veterans Health Administration National Drug File"

  -NCBI Taxonomy of Organisms

  -No 'Unreviewed' concepts!

- New version of Lexical Tools

  (Tutorial T25 Lexical Tools for UMLS Developers, Sunday 8:30 am)

- New version of Knowledge Source Server

Online Resources

WWW: http://www.nlm.nih.gov/research/umls/
http://umlsks.nlm.nih.gov
http://umlsinfo.nlm.nih.gov

E-mail: umls@nlm.nih.gov

umls-users listserv:

To subscribe to the listserv, send a message to

listserv@nlm.nih.gov

which includes the following line:

subscribe umls-users

To post a message to the umls-users listserv, AFTER subscribing, send email to:

umls-users@nlm.nih.gov

NATIONAL LIBRARY OF MEDICINE

Appendix

MRCON Strings

```
CUI      LAT TS  LUI     STT   SUI    STR                LRL
C0001403|ENG|P|L0001403|PF|S0010794|Addison's Disease|0|
C0001403|ENG|P|L0001403|VC|S0352253|ADDISON'S DISEASE|0|
C0001403|ENG|P|L0001403|VO|S0010792|Addison Disease|0|
C0001403|ENG|P|L0001403|VO|S0033587|Disease, Addison|0|
C0001403|ENG|P|L0001403|VO|S0469271|Addison's disease, NOS|3|
C0001403|ENG|S|L0278071|PF|S0352321|ADRENAL INSUFFICIENCY (ADDISON'S DISEASE)|0|
C0001403|ENG|S|L0278422|PF|S0352329|ADRENOCORTICAL INSUFFICIENCY, PRIMARY FAILURE|0|
C0001403|ENG|S|L0367999|PF|S0469267|Addison melanoderma|3|
C0001403|ENG|S|L0368000|PF|S0496840|Melasma addisonii|3|
C0001403|ENG|S|L0368398|PF|S0506528|Primary adrenal deficiency|3|
C0001403|ENG|S|L0373744|PF|S0471237|Asthenia pigmentosa|3|
C0001403|ENG|S|L0377831|PF|S0473611|Bronzed disease|3|
C0001403|ENG|S|L0494940|PF|S0718028|Primary adrenocortical insufficiency|3|
C0001403|ENG|s|L0494937|PF|S0718027|Primary adrenocortical insuff|3|
C0001403|FIN|P|L1510041|PF|S1805950|Addisonin tauti|3|
C0001403|FRE|S|L1272481|PF|S1514427|MALADIE D'ADDISON|2|
C0001403|GER|P|L1229627|PF|S1471573|Addison-Krankheit|3|
C0001403|GER|S|L1288823|PF|S1530769|Primaere Nebennierenrindeninsuffizienz|1|
C0001403|ITA|P|L1276837|PF|S1518783|Morbo di Addison|3|
C0001403|POR|P|L0324623|PF|S0432928|DOENCA DE ADDISON|2|
C0001403|RUS|P|L0889403|PF|S1093220|ADDISONOVA BOLEZN'|3|
C0001403|SPA|P|L0342625|PF|S0450930|ENFERMEDAD DE ADDISON|3|
[…]
```

MRSO Sources

```
CUI      LUI      SUI      SAB     TTY  SCD   SRL
C0001403|L0001403|S0010792|MSH2000|EN|D000224|0|
C0001403|L0001403|S0010794|MSH2000|MH|D000224|0|
C0001403|L0001403|S0010796|MSH2000|PM|D000224|0|
C0001403|L0001403|S0010796|PSY94|PT|00810|3|
C0001403|L0001403|S0219379|ICD91|IT|255.4|0|
C0001403|L0001403|S0220088|ICD91|IT|255.4|0|
C0001403|L0001403|S0220088|MSH2000|PM|D000224|0|
C0001403|L0001403|S0352252|CCPSS99|PT|0022753|3|
C0001403|L0001403|S0352252|DXP94|SY|NOCODE|0|
C0001403|L0001403|S0352253|CST95|GT|ADREN INSUFFIC|0|
C0001403|L0001403|S0352253|WHO97|IT|0410|2|
C0001403|L0001403|S0354372|AOD95|DE|0000005430|0|
C0001403|L0001403|S0354372|CSP98|PT|0060-3321|0|
C0001403|L0001403|S0354372|LCH90|PT|U000061|0|
C0001403|L0001403|S0354372|RCD99|PT|C1541|3|
C0001403|L0001403|S0354372|SNM2|SY|D-2332|3|
C0001403|L0001403|S0469271|SNMI98|PT|DB-70620|3|
C0001403|L0278071|S0352321|COS93|PT|U000087|0|
C0001403|L0278422|S0352329|DXP94|SY|NOCODE|0|
C0001403|L0367999|S0469267|SNMI98|SY|DB-70620|3|
C0001403|L0494937|S0718027|RCD99|AB|C1541|3|
C0001403|L0494940|S0718028|ICD10|PT|E27.1|3|
C0001403|L0494940|S0718028|RCD99|SY|C1541|3|
[…]
```

MRDEF Definitions

```
CUI      SAB     DEF
C0001403|MSH2000|A disease characterized by hypotension, weight loss, anorexia,
weakness, and sometimes a bronze-like melanotic hyperpigmentation of the skin. I
t is due to tuberculosis- or autoimmune-induced disease (hypofunction) of the ad
renal glands that results in deficiency of aldosterone and cortisol. In the abse
nce of replacement therapy, it is usually fatal.|
```

MRSTY Semantic Types

```
CUI      TUI  STY
C0001400|T040|Organism Function|
C0001403|T047|Disease or Syndrome|
C0001406|T083|Geographic Area|
C0001407|T114|Nucleic Acid, Nucleoside, or Nucleotide|
C0001407|T123|Biologically Active Substance|
```

MRATX Associated Expressions

```
CUI      SAB    REL     ATX
 Closed fracture of malar and maxillary bones, NOS
C0009045|MSH2000|B|<Zygomatic Fractures> OR <Maxillary Fractures>|
 Unilateral congenital dislocation of hip
C0009702|MSH2000|B|<Hip Dislocation, Congenital> AND <Femur Head>/<abnormalities>|
 Suture of bladder
C0010700|MSH2000|B|<Bladder>/<surgery>|
```

MRCXT Contexts

```
CUI      SUI      SAB    SCD    CXN CXL RNK   CXS              CUI2      HCD REL XC
C0001403|S0469271|SNMI98|DB-70620|1|ANC|1|SNOMED International|C0220967||||
C0001403|S0469271|SNMI98|DB-70620|1|ANC|2|DISEASES/DIAGNOSES|C0338067||||
C0001403|S0469271|SNMI98|DB-70620|1|ANC|3|DISEASES OF THE END. SYSTEM|C0014130||||
C0001403|S0469271|SNMI98|DB-70620|1|ANC|4|DISEASES OF THE ADRENAL GLANDS|C0001621||||
C0001403|S0469271|SNMI98|DB-70620|1|CCP||Addison's disease, NOS|C0001403|DB-70620|||

C0001403|S0718028|ICD10|E27.1|1|ANC|1|ICD, Tenth Revision (ICD-10)|C0391804||||
C0001403|S0718028|ICD10|E27.1|1|ANC|2|End., nutr. and metabolic diseases|C0694452||||
C0001403|S0718028|ICD10|E27.1|1|ANC|3|Disorders of other endocrine glands|C0178257||||
C0001403|S0718028|ICD10|E27.1|1|ANC|4|Other disorders of adrenal gland|C0494313||||
C0001403|S0718028|ICD10|E27.1|1|CCP||Primary adrenocortical insuff.|C0001403|E27.1|||

(* = C0001403|S0010794|MSH2000)
*|D000224|1|ANC|1|MeSH|C0220876||||
*|D000224|1|ANC|2|Diseases (MeSH Category)|C0012674|C|||
*|D000224|1|ANC|3|Endocrine Diseases|C0014130|C19|||
*|D000224|1|ANC|4|Adrenal Gland Diseases|C0001621|C19.53|isa||
*|D000224|1|ANC|5|Adrenal Gland Hypofunction|C0001623|C19.53.264|manifestation_of||
*|D000224|1|CCP||Addison's Disease|C0001403|C19.53.264.263|has_manifestation||
*|D000224|1|SIB||Adrenoleukodystrophy|C0001661|C19.53.264.270|has_manifestation||
*|D000224|1|SIB||Hypoaldosteronism|C0020595|C19.53.264.480|has_manifestation||
```

MRSAT String Attributes

```
CUI      LUI      SUI      SCD    ATN SAB     ATV
C0001403|L0001403|S0010792|D000224|EV|MSH2000|ADDISON DIS|
C0001403|L0001403|S0010794|D000224|AN|MSH2000|an autoimmune dis with adrenal hypofunction|
C0001403|L0001403|S0010794|D000224|DC|MSH2000|1|
C0001403|L0001403|S0010794|D000224|DE|MSH2000|ADDISONS DIS|
[…]
C0001403|L0001403|S0010794|D000224|M93|MSH2000|*120|
C0001403|L0001403|S0010794|D000224|M93|MSH2000|162|
C0001403|L0001403|S0010794|D000224|MED|MSH2000|*116|
C0001403|L0001403|S0010794|D000224|MED|MSH2000|167|
C0001403|L0001403|S0010794|D000224|MMR|MSH2000|19940628|
C0001403|L0001403|S0010794|D000224|MN|MSH2000|C19.53.264.263|
C0001403|L0001403|S0010794|D000224|MN|MSH2000|C20.111.163|
C0001403|L0001403|S0010794|D000224|TH|MSH2000|NLM (1966)|
C0001403|L0001403|S0352252|0022753|CCF|CCPSS99|44|
C0001403|L0001403|S0354372|C1541|RID|RCD99|Y41X1|
C0001403|L0001403|S0469271|DB-70620|SIC|SNMI98|255.4|
C0001403|L0367999|S0469267|DB-70620|SIC|SNMI98|255.4|
[…]
C0001403|L0494937|S0718027|C1541|RID|RCD99|Y41X2|
C0001403|L0494940|S0718028|C1541|RID|RCD99|Y41X2|
C0001403||||DA|MTH|19900930|
C0001403||||MR|MTH|20000101|
C0001403||||ST|MTH|R|
```

MRLO Locators

```
CUI      ISN                FR  UN         SUI    SNA SOUI
C0001403|MEDLINE(1990-1995)|228|*CITATIONS|S0010794|||
C0001403|MEDLINE(1996-Fall 1999)|116|*CITATIONS|S0010794|||
C0001403|DXPLAIN|||S0352252|||
C0001403|DXPLAIN|||S0352329|||
```

MRRANK Name Ranking

```
RANK SAB TTY SUPRES
0324|MTH|PN|N|
0323|MTH|MM|N|
0322|MSH2000|MH|N|
0321|MSH2000|HT|N|
0320|MSH2000|TQ|N|
0319|MSH2000|GQ|N|
0318|MSH2000|LQ|N|
0317|MSH2000|EP|N|
0316|MSH2000|EN|N|
0315|MSH2000|XQ|N|
0314|MSH2000|NM|N|
0313|DSM4|PT|N|
0312|DSM3R|PT|N|
0311|SNMI98|PT|N|
0310|SNMI98|PX|Y|
0309|SNMI98|HT|N|
0308|SNMI98|HX|Y|
0307|NDDF99|CD|N|
0306|NDDF99|IN|N|
0305|MDDB99|CD|N|
0304|MMX99|CD|N|
0303|MMX99|IN|N|
0302|RCDSA|PT|N|
[…]
```

MRREL Inter-concept Relationships

```
CUI1    REL CUI2   RELA  SAB  SL      MG
C0001403|AQ|C0205470||MSH2000|MSH2000||
C0001403|AQ|C0348026||MSH2000|MSH2000||
C0001403|CHD|C0271737||RCD99|RCD99||
C0001403|CHD|C0342477||RCD99|RCD99||
C0001403|PAR|C0001623|manifestation_of|MSH2000|MSH2000||
C0001403|PAR|C0004364|inverse_isa|MSH2000|MSH2000||
C0001403|PAR|C0405580||AOD95|AOD95||
C0001403|PAR|C0405580||RCD99|RCD99||
C0001403|PAR|C0494313||ICD10|ICD10||
C0001403|RB|C0001621||MTH|MTH||
C0001403|RB|C0004364||CSP98|MTH||
C0001403|RL|C0405580|mapped_from|SNMI98|SNMI98||
C0001403|RN|C0518933||MTH|MTH||
C0001403|RN|C0518934||MTH|MTH||
C0001403|RO|C0020615|clinically_associated_with|CCPSS99|CCPSS99||
C0001403|RO|C0041296||MTH|MTH||
C0001403|RO|C0085860|mapped_to|CSP98|CSP98||
C0001403|RO|C0151467|clinically_similar|RAM99|RAM99||
C0001403|RO|C0152889|associated_with|SNMI98|SNMI98||
C0001403|RO|C0405580|mapped_from|CST95|CST95||
C0001403|SIB|C0001661||MSH2000|MSH2000||
C0001403|SIB|C0002880||CSP98|CSP98||
[…]
```

MRCOC Co-occurrences

```
CUI1     CUI2    SOC COT COF COA
C0001403|C0000737|MBD|L|1|CO=1,DI=1|
C0001403|C0000833|MBD|L|1|DT=1|
C0001403|C0000833|MED|L|1|DT=1,MI=1,RA=1|
C0001403|C0001175|MBD|L|1|CO=1|
C0001403|C0001180|MBD|L|1|CO=1|
C0001403|C0001418|MBD|L|2|ET=2|
C0001403|C0001430|MED|L|1|BL=1,CO=1|
C0001403|C0001613|MBD|L|5|PP=2,CN=1,DI=1,HI=1,IM=1,SU=1|
C0001403|C0001613|MED|L|7|IM=4,ET=2,PP=2,BL=1,CL=1,PA=1|
C0001403|C0001614|MED|L|1|BL=1,CI=1|
C0001403|C0001617|MBD|L|1|BL=1|
C0001403|C0001618|MBD|L|1|IM=1|
C0001403|C0001618|MED|L|3|BL=2,CO=2,ET=1,PA=1|
C0001403|C0001621|MBD|L|10|ET=7,DI=3,PA=3,BL=1,CO=1,DT=1,PP=1|
C0001403|C0001621|MED|L|3|ET=3,DI=2|
C0001403|C0001623|MBD|L|7|DI=3,ET=2,PP=2,<>=1,CN=1,DT=1,IM=1,PA=1,TH=1|
C0001403|C0001623|MED|L|1|DI=1,ET=1|
C0001403|C0001624|MBD|L|10|ET=9,DI=2,DT=1,PA=1|
C0001403|C0001624|MED|L|3|DI=2,ET=2|
C0001403|C0001625|MBD|L|12|ET=4,CO=3,RA=3,SU=3,IM=2,BL=1,DT=1,EN=1,MI=1,PA=1,PP=1|
C0001403|C0001625|MED|L|7|IM=3,DI=2,PP=2,RA=2,BL=1,CO=1,ET=1,HI=1,PA=1,TH=1|
C0001403|C0001627|MBD|L|1|DT=1|
[…]
```

MRCON Suppressible synonyms

```
CUI      LAT TS  LUI     STT   SUI    STR                LRL
C0154009|ENG|P|L0180842|PF|S0245368|Benign neoplasm of prostate|0|
C0154009|ENG|P|L0180842|VO|S1650872|PROSTATE NEOPLASM BENIGN|3|
C0154009|ENG|P|L0180842|VO|S1912324|Neoplasm benign;prostate|3|
C0154009|ENG|P|L0180842|VO|S1933166|Neoplasm benign, prostate|3|
C0154009|ENG|S|L0524756|PF|S0599238|Benign tumor of prostate|3|
C0154009|ENG|S|L0524757|PF|S0599632|Benign tumour of prostate|3|
C0154009|ENG|S|L0524758|PF|S0598914|Benign prostatic tumor|3|
C0154009|ENG|S|L0524759|PF|S0598915|Benign prostatic tumour|3|
C0154009|ENG|s|L0033572|PF|S0999020|Prostate <3>|0|
C0154009|ENG|s|L0033572|VO|S0077252|Prostate|3|
C0154009|GER|P|L1258213|PF|S1500159|Gutartige Neubildung: Prostata|1|
```

SRDEF Basic information

```
RT  TUI  STY/RL  STN/RTN  DEF     EX     UN     NH     ABR     RIN
STY|T001|Organism|A1.1|Generally, a living individual, including all plants and
animals.|Homozygote; Radiation Chimera; Sporocyst|||||
STY|T002|Plant|A1.1.1|An organism having cellulose cell walls, growing by
synthesis of inorganic substances, generally distinguished by the presence of
chlorophyll, and lacking the power of locomotion.  Plant parts are included here
as well.|Pollen; Potatoes; Vegetables|||||
STY|T003|Alga|A1.1.1.1|A chiefly aquatic plant that contains chlorophyll, but does
not form embryos during development and lacks vascular tissue.|Chlorella;
Laminaria; Seaweed|||||
STY|T004|Fungus|A1.1.2|A eukaryotic organism characterized by the absence of
chlorophyll and the presence of a rigid cell wall.  Included here are both slime
molds and true fungi such as yeasts, molds, mildews, and mushrooms.|Aspergillus
clavatus; Blastomyces; Helminthosporium; Neurospora|||||
[…]
RL|T132|physically_related_to|R1|Related by virtue of some physical attribute or
characteristic.||||PR|physically_related_to|
RL|T133|part_of|R1.1|Composes, with one or more other physical units, some larger
whole.  This includes component of, division of, portion of, fragment of, section
of, and layer of.||||PT|has_part|
[…]
RL|T186|isa|H|The basic hierarchical link in the Network.  If one item "isa"
another item then the first item is more specific in meaning than the second
item.||||IS|inverse_isa|
[…]
```

SRSTR Structure

```
STY/RL              RL      STY/RL                                  LS
Biologic Function|affects|Organism|D|
Biologic Function|isa|Natural Phenomenon or Process|D|
Biologic Function|process_of|Organism|D|
Biologic Function|produces|Biologically Active Substance|D|
Biologic Function|produces|Body Substance|D|
[…]
Disease or Syndrome|conceptually_related_to|Experimental Model of Disease|DNI|
Disease or Syndrome|isa|Pathologic Function|D|
Disease or Syndrome|produces|Tissue|D|
[…]
Medical Device|isa|Manufactured Object|D|
Medical Device|prevents|Injury or Poisoning|D|
Medical Device|prevents|Pathologic Function|D|
Medical Device|treats|Anatomical Abnormality|D|
Medical Device|treats|Injury or Poisoning|D|
Medical Device|treats|Pathologic Function|D|
Medical Device|treats|Sign or Symptom|D|
[…]
Mental Process|process_of|Plant|B|  blocks  Biologic Function|process_of|Organism|D|
[…]
part_of|isa|physically_related_to|D|
[…]
```

SRSTRE2 Structure (expanded)

```
STY                  RL          STY
Disease or Syndrome|isa|Pathologic Function|
Disease or Syndrome|isa|Biologic Function|
Disease or Syndrome|isa|Natural Phen. or Pr.|
Disease or Syndrome|isa|Phenomenon or Process|
Disease or Syndrome|isa|Event|
Disease or Syndrome|affects|Alga|
Disease or Syndrome|affects|Amphibian|
Disease or Syndrome|affects|Animal|
Disease or Syndrome|affects|Archaeon|
Disease or Syndrome|affects|Bacterium|
Disease or Syndrome|affects|Biologic Function|
Disease or Syndrome|affects|Bird|
Disease or Syndrome|affects|Cell Function|
Disease or Syndrome|affects|Cell or Molecular Dysfunction|
[…]
```

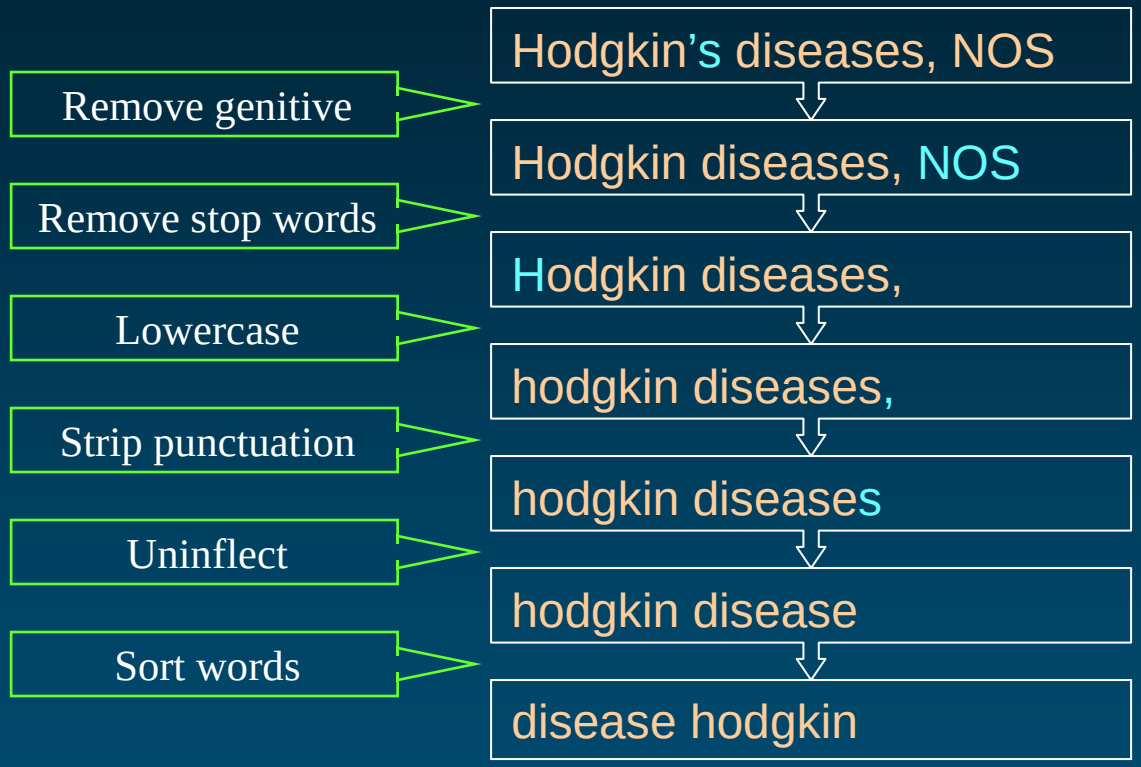
```
Pathologic Function |isa|Biologic Function|
Biologic Function|isa|Natural Phen. or Process|
Natural Phen. or Process|isa|Phen. or Process|
Phenomenon or Process|isa|Event|
```

from `Biologic Function|affects|Organism|D|`

Normalization Example

Hodgkin Disease
HODGKINS DISEASE
Hodgkin's Disease
Disease, Hodgkin's
Hodgkin's, disease
HODGKIN'S DISEASE
Hodgkin's disease
Hodgkins Disease
Hodgkin's disease NOS
Hodgkin's disease, NOS
Disease, Hodgkins
Diseases, Hodgkins
Hodgkins Diseases
Hodgkins disease
hodgkin's disease
Disease, Hodgkin

Normalization



Addison's Disease: Co-occurring concepts

```
25  Autoimmune Diseases
21  Autoantibodies
20  Hydrocortisone
19  Adrenal Glands
16  Steroid 21-Monooxygenase
13  Adrenal Gland Diseases
13  Adrenal Gland Neoplasms
12  Polyendocrinopathies, Autoimmune
12  Adrenal Cortex
11  Tuberculosis, Endocrine
10  Corticotropin
10  Glucocorticoids
 9  Diabetes Mellitus, Insulin-Dependent
 8  Thyroiditis, Autoimmune
 8  Tuberculosis
 8  Hypothyroidism
 8  Adrenal gland hypofunction
 8  Autoantigens
 8  Adrenoleukodystrophy
[…]
 1  Circadian Rhythm
[…]
```